Identification of regulatory RNA elements based on structural conservation

Dem Fachbereich Biologie der
Technischen Universität Darmstadt
zur Erlangung des akademischen Grades eines
Dr. rer. nat

eingereichte Dissertation

von
Herr M. Sc. Johannes Braun
aus
Landau i. d. Pfalz

1. Referentin: Prof. Dr. Beatrix Süss
2. Referent: Prof. Dr. Viktor Stein

Datum der Einreichung: 04.12.2018
Tag der mündlichen Prüfung: 24.01.2019
Darmstadt 2019

Braun, Johannes: Identification of regulatory RNA elements based on structural conservation
Darmstadt,Technische Universität Darmstadt,
Jahr der Veröffentlichung der Dissertationauf TUprints: 2019
Tag der mündlichen Prüfung: 24.01.2019

Für meine Mädels.

The posttranscriptional regulation of gene expression determines the amount of a protein produced from a specific mRNA. All stages of the mRNA life cycle are post-transcriptionally regulated. The untranslated regions (UTRs) of mRNA play an important role in this process. UTRs encode *cis*-regulatory elements that interact with *trans*-acting factors such as RNA binding proteins or non-coding RNAs. These interactions are based on the specific recognition of sequence or structured motifs. Similar to the conservation of linear sequences, the conservation of secondary structures can be an indicative of a functional *cis*-regulatory element.

In the first part of my doctoral thesis, I identified structurally conserved regulatory elements in 3'UTRs of mRNAs. For this, I performed reporter gene assays with bioinformatically predicted structurally conserved RNA elements and discovered a regulatory element in the 3'UTR of the *UCP3* (uncoupling protein 3) mRNA. UCP3 is a protein of the inner mitochiondrial membrane and is associated with the development of *diabetes melitus* type 2 (T2DM). Through sequence and structural analysis, I discovered that the element has an active conformation that consists of two short RNA stem-loops. In further experiments, I was able to confirm that the presence of both RNA stem-loops is necessary for efficient regulation. Furthermore, I showed that the reduction of reporter gene expression is caused by a reduction of the mRNA half-life. The prediction of conserved RNA structures thus provides a powerful tool for the *de novo* identification of *cis*-regulatory elements.

In the second part of my doctoral thesis, I characterized the regulatory element from the 3'UTR of *UCP3* in detail. First, I performed RNA affinity purification to identify proteins specifically associated with the *UCP3* element by mass spectrometry. This allowed me to show that the proteins Roquin-1 and Roquin-2 bind to the RNA stem loops of the UCP3 element. Furthermore, I showed that endogenous *UCP3* is regulated by Roquin. Roquin proteins bind to constitutive (CDEs) and alternative (ADEs) decay elements and induce the rapid degradation of mRNAs of genes that play an important role in the immune response. Binding studies with the Roquin-1 ROQ domain showed that Roquin binds with significantly higher affinity to the *UCP3* element, when both CDEs are present. Both CDEs in the *UCP3* element do not correspond to the previously suggested CDE consensus. Performing a detailed mutational analysis, I revised the CDE consensus. With my data >160 new CDE- and 19 new ADE-coding mRNAs could be identified. Furthermore, I confirmed new CDE/ADE-containing mRNAs as targets of Roquin. Interestingly, I was able to show that not only the expression of CDE-encoding mRNAs, but also regulation by Roquin is cell type dependent. In conclusion, I have extended Roquin's role in the posttranscriptional regulation of gene expression and suggest that its role is not limited to the regulation of the immune response.

Die posttranskriptionelle Regulation der Genexpression bestimmt maßgeblich die Menge eines Proteins, welches von einer bestimmten mRNA hergestellt wird. Alle Prozesse im Lebenszyklus einer mRNA werden posttranskriptionell reguliert. Eine besondere Rolle spielen dabei die nicht translatierten Bereiche (UTRs) der mRNA. UTRs kodieren *cis*-regulatorische Elemente welche mit *trans*-agierende Faktoren wie RNA Bindeproteinen oder nicht-kodierenden RNAs interagieren. Diese Interaktionen basieren auf der spezifischen Erkennung von Sequenz- oder Strukturmotifen. Ähnlich wie die Konservierung von Sequenzen, kann auch die Konservierung von Sekundärstrukturen Hinweis auf ein funktionales *cis*-regulatorisches Element sein.

Im ersten Teil meiner Doktorarbeit identifizierte ich strukturkonservierte, regulatorische Elemente in 3'UTRs der mRNA. Dazu führte ich Reportergenanalysen mit bioinformatisch vorhergesagten strukturkonservierten RNA Elementen durch und entdeckte ein regulatorisches Element im 3'UTR der *UCP3* (uncoupling protein 3) mRNA. UCP3 ist ein mitochondriales Membranprotein und mit der Entstehung von *Diabetes Melitus* Typ 2 assoziiert. Durch Sequenz- und Strukturanalyse fand ich heraus, dass das Element eine aktive Konformation mit zwei kurzen RNA Stammschleifen einnimmt. In weiteren Experimenten konnte ich bestätigen, dass die Anwesenheit beider RNA Stammschleifen notwendig für die Regulation ist. Darüber hinaus zeigte ich, dass die Reduktion der Reportergenexpression durch eine drastische Verringerung der mRNA Halbwertszeit verursacht wird. Die Vorhersage konservierter RNA Strukturen, stellt somit ein effizientes Werkzeug zur *de novo* Identifikation *cis*-regulatorischer Elemente dar.

Im zweiten Teil meiner Doktorarbeit untersuchte ich das regulatorische Element aus dem 3'UTR von *UCP3* im Detail. Zunächst führte ich RNA Affinitätsaufreinigung durch um Proteine, die speziefisch mit dem *UCP3* Element interagieren, mittels Massenspektromentrie zu identifizieren. Dadurch konnte ich zeigen, dass die Proteine Roquin-1 und Roquin-2 an die RNA Stammschleifen des *UCP3* Elements binden. Darüber hinaus zeigte ich, dass endogenes *UCP3* durch Roquin reguliert wird. Roquin Proteine binden an konstitutive (CDEs) und alternative (ADEs) Abbauelemente und induzieren eine schnellen Abbau der mRNA von Genen die eine wichtige Rolle bei der Immunantwort spielen. Bindungstudien mit der Roquin-1 ROQ Domäne zeigten, dass Roquin mit deutlich höherer Affinität an das *UCP3* Element bindet, wenn beide CDEs vohanden sind. Beide CDEs im *UCP3* Element entspechen nicht dem bisherigen Consesus für CDEs. Mit Hilfe einer Mutationsstudie revidierte ich den CDE Konsesus. Mit meinen Daten konnten >160 neue CDE- und 19 neue ADE-kodierende mRNAs identifiziert werden. Drüber hinaus bestätigte ich neue CDE/ADE-kodierende mRNAs als Ziele von Roquin. Interessanterweise konnte ich zeigen, dass nicht nur die Expression CDE-kodierender mRNAs, sondern auch die Regulation durch Roquin, Zelltyp abhängig ist. Durch meine Arbeit konnte die Rolle von Roquin in der posttranskriptionellen Regulation der Genexpression erweitert werden und und gezeigt werden, dass die Regulation durch Roquin nicht auf Gene der Immunantwort beschränkt ist.

Contents

List of Figures

1 Introduction

The regulation of gene expression in eukaryotes is complex and tightly controlled, and is commonly described as the process by which the genetic code is used to direct protein synthesis. The abundance of a protein is not only regulated at transcriptional level, but also at several post-transcriptional levels. Post-transcriptional regulation (PTR) is achieved at all steps in the complex life of a messenger RNA (mRNA), from its birth over pre-mRNA splicing, through mRNA export to translation. An ensemble of RNA-binding proteins (RPB) as well as non-coding RNAs (ncRNA) can interact with a given mRNA, forming a dynamic messenger ribonucleoprotein particle (mRNP) and thereby determine the fate of an mRNA from synthesis to decay [1]. The untranslated regions (UTRs) of an mRNA play a particularly important role, as they encode *cis*-regulatory elements that interact with *trans*-acting factors, i.e. RBPs and ncRNAs. While the 5'UTR of an mRNA generally regulates translation, the function of 3'UTRs is more divers. Regulatory elements in 3'UTRs determine protein abundance by the regulation of mRNA stability, mRNA localization and translation efficiency. Next generation sequencing methods provided us not only with a landscape of 3'UTRs, but also with the discovery of a plethora of RBP and ncRNA binding sites that directly contribute to the regulation of the respective mRNA.

This doctoral thesis focuses on an important part in understanding RNA-based regulation that intends to understand how regulatory mRNA structures contribute to PTR. A variety of concepts emerged in this field to understand *cis*-regulatory elements and their interplay with *trans*-acting factors. But, while manifold methods were developed for the prediction and identification of linear sequence motifs over the last decade, structured RNA elements and their contribution to PTR are more difficult to assess and therefore less studied.

1.1 Post-transcriptional Regulation

The regulation of gene expression can be divided into two broad categories: transcriptional and post-transcriptional events. Transcriptional control concerns all processes that determine the amount of transcript produced from a given gene. PTR on the other hand, includes all regulatory events that directly influence mRNA and thereby determine protein abundance.

In the process of transcription, RNA polymerases synthesize pre-mRNA molecules based on DNA matrices. This process predominantly relies on the accessibility of a

given genomic locus by the transcription machinery. Accessibility is achieved by dynamic remodeling of higher order genome structures. The primary structure of eukaryotic chromosomes consist of nucleosomes, a stable complex that constitutes of the double-stranded DNA and histone proteins. Contiguous nucleosomes are separated by short sections of linker DNA and yield the classic bead-on-a-string structure. These nucleosomes are then organized in a higher order structure the chromatin, which involves folding of the beads-on-a-string motif into a ∼30 nm thick fiber [2]. In order to provide access to the DNA, the chromatin structure can be altered by modifying the strength of the DNA-histone interaction. This is achieved by a number of post-translational modifications, i.e. acetylation, methylation, phosphorylation, ubiquitylation and sumoylation, of specific amino acid residues of the histones, resulting in open or closed chromatin [3] [4]. Additionally, ATP-dependent chromatin remodellers are involved in sliding, evicting and restructuring of nucleosomes [3]. Most protein-coding genes in eukaryotes are transcribed by RNA polymerase II. This process is dependent on regulatory proteins, transcription factors, which are mandatory for the correct recruitment and binding of the polymerase to the DNA. These factors recognize the core promoter of a gene, but also nearby regulatory elements and distal elements in *cis*, which can be enhancers, silencers, insulators or locus control regions, which function either to enhance or repress transcription [5]. Therefore transcription factors significantly contribute to selective gene expression in cells.

In contrast to prokaryotes where transcription is directly coupled to translation, those two processes are spatially separated in eukaryotes as result of compartmentalization: transcription takes place in the nucleus, whereas translation takes place in the cytoplasm. The uncoupling of transcription and translation provides the cell with extensive opportunities to regulate the amount of protein that is synthesized, at the mRNA level. PTR is evident throughout the different stages of an mRNA's life [6] [7]. A schematic representation of the eukaryotic mRNA life cycle is depicted in **Fig. 1.1**.

As soon as mRNA precursors are transcribed, the nascent pre-mRNA associates with proteins that are necessary for further processing. Initially, the nascent pre-mRNA is modified at its 5' end by addition of a 7-methylguanosine in a co-transcriptional manner. The 5' end associates with specific protein factors, i.e the cap-binding complex (CBC) that also facilitates splicing, 3' end formation and even export as it diffuses with bound mRNAs to nuclear pores [9]. After the export into the cytoplasm the CBC is exchanged for the translation initiation factor EIF4E [8]. During transcription the non-coding intronic sequences of an mRNA are removed by the process of splicing. This is facilitated by a multi-protein complex, the spliceosome. Alternative splicing vastly increases protein diversity of eukaryotic genomes. Alternative splicing can also label an mRNA for specific degradation in the cytoplasm [10]. The 3' end of the transcript encodes a polyadenylation signal sequence -AAUAAA- that is recognized by cleavage and polyadenylation specificity factor 6 (CPSF6), which mediates cleavage of the nascent pre-mRNA and addition of a 150-200 nt polyadenine (A) tail to the transcript. The poly(A) tail of the mRNA associates with poly(A) binding proteins (PABPs) and me-

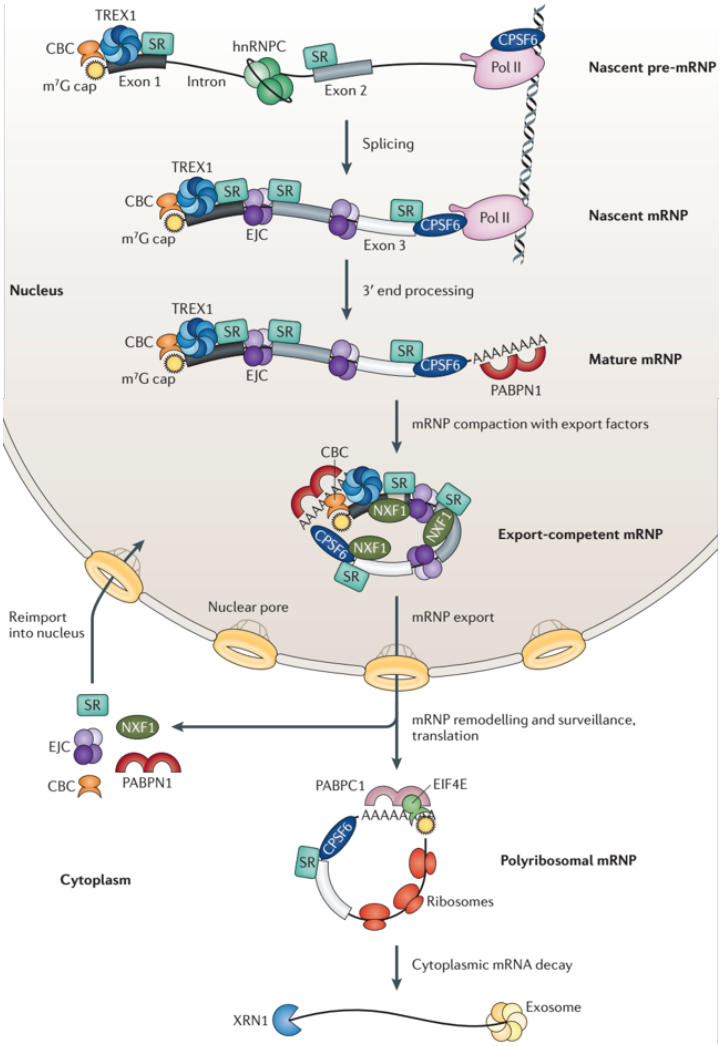

Figure 1.1: The complex life of mRNA. Illustration of the mRNA life cycle from transcription to translation to decay. Important RBPs that are crucial at various steps are shown. Most RPBs that associate with the mRNA are replaced during transcription. For details refer to the main text. CBC, cap-binding complex; m7G, 7-methylguanosine cap; EIF4E, eukaryotic initiation factor 4E; CPSF6, cleavage and polyadenylation specificity factor 6; EJC, exon junction complex; hnRNPC, heterogeneous nuclear ribonucleoprotein C; NXF1, nuclear export factor 1; PABPC1, cytoplasmic poly(A) binding protein 1; PABPN1, nuclear poly(A)-binding protein 1; Pol II, RNA polymerase II; TREX1, transcription export complex 1; XRN1, 5'-3' exoribonuclease 1. Adopted from [8].

diates circularization of the mRNA by interaction with 5' end interacting proteins and thereby forms the mRNP. mRNA circularization increases mRNA stability in the cytoplasm as well as the translation efficiency. The dynamic composition of an mRNP in the cytoplasm dictates the subsequent fate of an mRNA by regulating its localization, stability and translation efficiency [6] [8]. However, at the end of the mRNA life cycle lies always the degradation by nucleases. The degradation of an mRNA can be either dependent on the removal of 5' cap or the 3' poly(A) tail and is extensively regulated by *cis*-regulatory element within the UTRs of an mRNA.

1.1.1 Pre-mRNA Splicing

In eukaryotic genes, the protein-coding sequence is interrupted by intervening sequences (introns), which must be removed by splicing. This process is catalyzed by a dynamic ribonucleoprotein complex, the spliceosome, mostly in a co-transcriptional manner, but it can also occur post-transcriptionally [11] [12]. Co-transcriptional splicing allows for the regulation by diverse transcription-dependent mechanisms, whereas post-transcriptional splicing could create additional layers of regulation by coupling splicing with downstream events like mRNA export [13]. Nuclear pre-mRNA splicing contains two simple transesterification reactions that involve reactive groups from three regions in the pre-mRNA. Self-splicing group II introns fold into a complex secondary and tertiary structure, which juxtaposes the reactive groups in an intron and thereby forms an active site. In contrast, bringing reactive groups in a pre-mRNA intron into close proximity is dependent on the *trans*-acting factors comprising the spliceosome [14]. The major spliceosome consists of five small nucleolar ribonucleoproteins (snRNP) that dynamically interact with three specific regions to exert pre-mRNA splicing. These important features are: the 5' splice site (SS), the 3'SS and the branch point (BP) with its adjacent poly(Y)-tract. Usually introns encode a 5'SS that starts with a GU dinucleotide and a 3'SS that ends with an AG dinucleotide. The third important feature is the BP, which is located 18-40 nt upstream of the 3'SS and characteristically displays a -YNYURAY- motif.

In the first step of the splicing reaction the 2'-OH of the BP adenosine performs a nucleophilic attack on the phosphodiester backbone at the 5'SS resulting in a free 5' exon and an intron-3' exon lariat. In the second step, the 5' exon 3'-OH attacks the phosphodiester backbone at the 3'SS, leading to exon ligation and lariat intron excision. In order to bring all relevant protein and RNA components together at the pre-mRNA the spliceosome comprises a large number of subunits, to accommodate for all relevant reaction steps and their regulation. The assembly of the spliceosome on the pre-mRNA is orchestrated by subsequent interactions of the spliceosomal snRNPs. Initially, the U1 snRNP recognizes the 5'SS by base pairing interactions and the factors SF1/mBBP and U2AF interact with the BP and poly(Y)-tract. Then, the U2 snRNP forms a stable complex with the BP and thereby forms the pre-spliceosome (also A-complex). A pre-assembled complex of the snRNP U4/U6 and U5 than is recruited to the A-complex resulting in a pre-catalytic B-complex, which is subsequently rearranged,

leading to release of U1 and U4 and results in the activated spliceosome. After catalytic activation by auxiliary factors the first transesterification can be catalyzed, which yields the complex required for the second step of the splicing reaction [15]. In an energy consuming step the spliceosome is than resolved and its components reused in subsequent rounds of splicing reactions.

There are different pathways for early stage spliceosome assembly in metazoans. As most mammalian pre-mRNAs are composed of ∼120 nt long exons and introns which vary in length from hundreds to thousands of nucleotides, splicing complexes initially define the exons. In this process, the U1 snRNP binding to the 5'SS downstream of an exon promotes U2AF association to the 3' upstream poly(Y)-tract which leads to the recruitment of the U2 snRNP to the upstream BP. This exon defined complex is stabilized by interaction of serin/argine-rich (SR) proteins with splicing enhancer sequences, that establish an exon-spanning network of protein-protein interactions [15].

1.1.2 Alternative pre-mRNA Splicing

Alternative splicing (AS) is an essential and ubiquitous mechanism that plays an important role in almost every biological process over the entire life span of an organism from its birth to its death [16]. AS post-transcriptionally regulates gene expression beyond turning genes "on" and "off". In general, AS refers to any form of splicing that variably includes or excludes exon, introns or fractions thereof, in a set of transcripts from a given gene [17]. Thereby, AS contributes by large margins to eukaryotic protein diversity. About 95% of human genes undergo AS, making it the most prominent RNA processing mechanism that generates mRNA complexity. The variable inclusion or exclusion of sequences in the protein-coding region of a gene, typically leads to different protein isoforms that contain specific peptide domains or yields mRNAs harboring premature stop codons (PTCs), which trigger mRNA degradation by nonsense-mediated decay (NMD). In contrast, AS of 5' or 3'UTRs can alter the presence of *cis*-regulatory elements within those regions and thereby influences mRNA localization, stability and translation efficiency. Taken together, this highlights the fundamental role of AS as a mechanism that is critical for the proper function of cells [18] [16]. The importance of AS is further highlighted by the fact that corruption of the process is associated with numerous pathologies like cancer and improper immune responses.

There are four major AS events. Almost 40% of AS events in higher eukaryotes are exon skipping events, followed by alternative 5'SS and 3'SS selection accounting for 7.9% and 18.4% of AS events, respectively. The rarest type of AS event is intron retention which, only accounts for less than 5% of events in vertebrates and invertebrates and is most prevalent in UTRs. In addition to that, less frequent events occur that lead to transcripts with mutually exclusive exons, alternative promoters and alternative polyadenylation. All of the four main splicing events can occur in the protein coding sequence as well as in UTRs [19].

Alternative SS usage by the spliceosome is determined by the interaction of addi-

tional proteins with specific pre-mRNA sequences, so-called intronic and exonic splicing enhancer or silencer sequences, under particular cellular conditions. SR proteins are splicing regulators that are best known to recognize exonic splicing enhancers and promote exon inclusion. On the other hand, heterogeneous ribonucleoproteins (hnRNPs) often mediate exon skipping by interaction with exonic splicing silencers. AS therefore reflects the occupancy and interaction of these regulatory elements with different proteins that are present in a specific cellular scenario [20].

1.1.3 mRNA Surveillance

In order to ensure cell function and viability, gene expression must maintain a high level of fidelity. Errors during transcription or RNA processing can lead to non-functional transcripts, encoding truncated proteins that are potentially deleterious for a cell. At the mRNA level, several quality control mechanisms evolved that operate during translation in order to protect the cell from aberrant mRNAs: non-stop decay (NSD), no-go decay (NGD) and NMD. NSD targets transcripts for degradation that do not contain a stop codon. NGD targets mRNAs with ribosomes stalled during elongation. NMD promotes degradation of mRNAs that undergo premature translation termination by encoding a premature stop codon [21] [22].

The best described RNA surveillance mechanism in eukaryotes is NMD, which provides a powerful means of not only controlling mRNA quality, but regulation of gene expression as well. The fact, that 30% of all mutations involved in human diseases lead to a PTC, highlight the biological importance of NMD [22]. NMD targets at least two different sets of mRNAs. The first important group of NMD targets are mRNA transcripts that harbor a PTC that truncates their open reading frame. In general, a stop codon located $\sim$55 nt upstream of an exon-junction complex triggers NMD [23]. These transcripts arise from different origins. Mutation at the DNA level can introduce PTCs by direct nonsense mutation or by frame-shift inducing deletions or insertions. More importantly, PTC containing mRNAs can be produced by AS. Computational predictions propose, that one third of alternatively spliced mRNAs harbor a PTC, highlighting the coupling of NMD to AS. Various studies revealed how AS events exploit NMD as a means to regulate the abundance of transcripts under certain cellular conditions. This type of regulation therefore has been termed AS-NMD or RUST (regulated unproductive splicing and translation) and is found in the regulation of many splicing regulators, like SR proteins and hnRNPs, indicating its role as important feedback regulator of splicing. Because NMD sensitive mRNA isoforms are not translated into a protein, AS-NMD allows for the control of overall protein abundance by switching between transcripts that are translated or degraded. Approximately 10-15% of all human genes are regulated by the AS-NMD mechanism [24].

The second group comprises physiological transcripts that encode full-length proteins. The role of NMD as regulator of physiological mRNAs was highlighted in transcriptome profiling studies that revealed, that NMD controls the abundance of 3-10% of cellular

genes. Features that render physiological mRNAs sensitive for NMD include introns located in the 3'UTR, upstream open reading frames (uORFs), programmed frame-shifts and long 3'UTRs. An interesting case for NMD substrates are UGA codon bearing transcripts, which can encode the amino acid selenocystein. When selenium concentration is high UGA encodes selenocystein, if selenium concentration is low UGA is interpreted as PTC [24].

1.1.4 mRNA Decay

In general, eukaryotic mRNAs have 5' cap structures and 3' poly(A)-tails. These features are essential for the formation of the mRNP, which protects the mRNA from decay by exonucleases and is necessary for efficient translation. However, this state is dynamic and transient. The rate limiting steps of mRNA decay are the shortening of the poly(A)-tail of an mRNA and removal of the 5' cap structure. In eukaryotes, two distinct pathways for mRNA decay exist. After deadenylation mRNAs can be degraded in either 3'→5' direction by the exosome or in 5'→3' by the nuclease Xrn1 [25].

Initially, the the main deadenylation machinery, a complex consisting of Ccr4–Not in yeast or PARN (poly(A)-specific ribonuclease) and PAN2/PAN3 (polyA-specific ribonuclease subunit) in metazoans, facilitates the loss of PABP by shortening of the poly(A)-tail which leads to the disruption of the circular mRNP structure. mRNA linearization enables the association of additional factors to both ends of an mRNA molecule. One mechanism for cytoplasmatic mRNA turnover is the deadenylation-dependent 5'→3' degradation. This process is limited by the rate of decapping. Decapping of the accessible 5' end is executed by the Dcp1p/Dcp2p decapping complex in conjunction with the activity of several accessory factors. These include the heptameric Lsm1-7p complex, the Dhh1p RNA helicase (also known as RCK/p54 or Me31B), and Pat1p, which are proposed to facilitate the association of the decapping complex. This process is further modulated by enhancing factors, the Edc1, Edc2, and Edc3 (Lsm16) proteins. Interestingly Dhh1p and Pat1p which stimulate the decapping complex, also serve as translational repressors. After decapping, the mRNA is subsequently degraded by the exonuclease Xrn1 [25] [21].

The alternative post-deadenylation mRNA decay pathway is the 3'→5' degradation by the multi-subunit exosome complex. A nine subunit containing catalytic inactive core exosome complex is present in both the nucleus and the cytoplasm. For catalytic activity different additional factors associate with the exosome in each compartment. In addition to the nine core exosome subunits, catalytically active exosomes need the nuclease factor Dis3p/Rrp44p and the 3'→5' exonuclease Rrp6p/Ski7p in the nuclear or cytoplasmic exosome, respectively [26].

mRNAs that undergo degradation have been observed in two distinct translational states. On the one hand, translationally inactive mRNAs localize to P-bodies, distinct cytoplasmatic foci that are enriched in components of the mRNA degradation machinery, where they undergo degradation. On the other hand, actively translated mRNAs present

in polyribosome fractions can be degraded. Deletion of the key P-body genes Edc3 and Lsm4 did not impair mRNA decay of individual mRNAs suggesting that localization to P-bodies is not obligatory for mRNA degradation. This is further highlighted by experiments that showed that polysome fractions contain deadenylated, decapped, and partially exonucleolytically degraded mRNAs, and the finding that mRNA decay factors cofractionate with ribosomes [21].

1.1.5 Non-coding RNAs in Gene Regulation

High-throughput transcriptome studies revealed that ∼90% of the human genome is transcribed. Yet, only 1-2% of the genome encodes for proteins. Thus, the vast majority of transcripts are non-coding (nc) RNAs, which have regulatory functions [27]. Evidence supports that ncRNAs play a role in the response to stress and environmental stimuli as well as in developmental processes. NcRNAs can be divided into two main groups: infrastructural ncRNAs and regulatory ncRNAs. The first group comprises ribosomal, transfer, small nuclear and nucleolar RNAs. These RNAs function as scaffolds in dynamic RNA-protein complexes that are involved in every step of an mRNA's life. The second group of ncRNAs comprises regulatory ncRNAs: microRNAs (miRs), Piwi-interacting RNAs (piRNAs), small interfering RNAs (siRNAs), and long non-coding RNAs (lncRNAs). PiRNAs are a group of 24–31 nt long ncRNAs that form complexes with Piwi proteins of the Argonaute protein family and thereby suppress transposon activity during germline development. SiRNAs are linear, perfectly base-paired dsRNA, which is processed to a 21-24 nt long siRNA by Dicer and mediates post-transcriptional silencing when loaded onto Argonaute proteins. The siRNA mode of action is similar to that of miRs (see below). LncRNAs represent the largest group of non-coding transcripts and are generally >200 nt long. The majority of lncRNAs are transcribed as complex networks of sense and antisense transcripts regarding their respective genomic locus. Antisense transcripts have been shown to modulate mRNA stability, e. g. of hypoxia inducible factor alpha (HIF-1a), nitric-oxide synthase (eNOS) and tyrosine kinase with immunoglobulin-like and EGF-like domains 1 (tie-1), which highlights their importance in PTR [27].

Probably the most elaborately studied group of regulatory ncRNAs are miRs. They comprise a large family of ∼21 nt long ncRNAs that are present in most eukaryotic organisms and are key-regulators of gene expression [28]. The first the miRs, lin-4 and let-7, were identified in *C. elegans* screenings for genes that control larva development [29]. Typically, miRs are associated with the repression of protein-coding genes, but rare examples were miRs may lead to the induction of a specific gene under certain conditions also exist [30] [31]. The widespread involvement of miRs in PTR is substantiated by >2,500 miRs that have already been discovered in humans (collected in the miRBase 20.0 database) [32], of which >600 have been experimentally confirmed [33]. Approximately 50-60% of all protein-coding genes are regulated by miRs [34] [35]. MiRs

regulate gene expression by modulation of mRNA stability or translation efficiency of their target gene. They often interact with the 3'UTRs of target transcripts [36]. MiRs mature by a complex biogenesis, which itself is highly regulated [37] [38] [39]. MiR genes are generally encoded in intergenic regions or introns. They are typically transcribed by RNA polymerase II to produce the primary (pri)-miR transcript that harbors an RNA hairpin structure that co-transcriptionally associates with the multi-protein Microprocessor complex. The Microprocessor contains the RNase III Drosha and the RBP DGCR8 and crops the pri-miR to a ∼60 nt hairpin-structured pre-miR. Alternatively, pre-miRNA-like introns, called mirtrons can produce pre-miRs by splicing, independent of the Drosha processing. Next, the pre-miR is actively transported from the nucleus to the cytoplasm by Exportin-5 (Exp5), where the proteins Dicer, TRBP and PACT assemble on the pre-miR and cleave it into a ∼21 nt miR duplex. This duplex is loaded onto an RNP complex called RISC (RNA-induced silencing complex), which facilitates degradation of the passenger strand. The miR-RISC complex then associates with partially complementary sequences in an mRNA, depending on the accessibility of the miR binding site and the occurrence of the minimal perfect Watson–Crick base pairing nucleotides of the seed region between miR and target mRNA [28].

1.1.6 RNA-binding Proteins

Along with miRs, PTR of gene expression is achieved by the dynamic interactions of RNAs with RBPs that assemble into an RNP. Generally RBPs contain one or multiple RNA-binding domains [40], by which they recognize their targets and orchestrate all aspects of the mRNA life cycle from production over processing to mRNA function. Key components of mRNPs, like the cytoplasmic cap–binding proteins, PABP and hnRNPs were initially identified in the 1970s [41]. More than 1500 RBPs were identified since, most of them by RNA interactome studies over the last decade [33] [42] [43]. Deregulation or dysfunction of RBPs is often associated with disease, highlighting the importance of RBPs in cellular homeostasis [44] [43]. RBPs change the fate of an mRNA upon binding by modulating RNA processing, modification, stability, translation and localization. Inversely, RBP fate can be changed by RNA binding by altering RBP function, facilitating protein-protein interaction or modulation of stability and spatio-temporal distribution (**Fig. 1.2**) [45] [46]. Approximately 3-11% of eukaryotic genomes encode RBPs. Most RBPs contain a set of well defined RNA binding domains (RBD) such as the RNA recognition motif (RRM), K-homology domain (KH), CCCH zinc finger domain (ZnF) and double-stranded RNA binding motifs, like the DEAD box helicase domain (DHX) [47]. By the modular combination of RBDs and their cooperative function, RBPs build up affinity and specificity for their respective RNA targets [44]. In addition, an increasing number of RBPs lack classical RBDs at all and contain protein domains with dual functions that can perform enzymatic activities and RNA binding or contain protein folds of unknown function or lacking any defined tertiary structure in the unbound state [44]. Binding affinity and specificity is also influenced by composite RNA motifs that

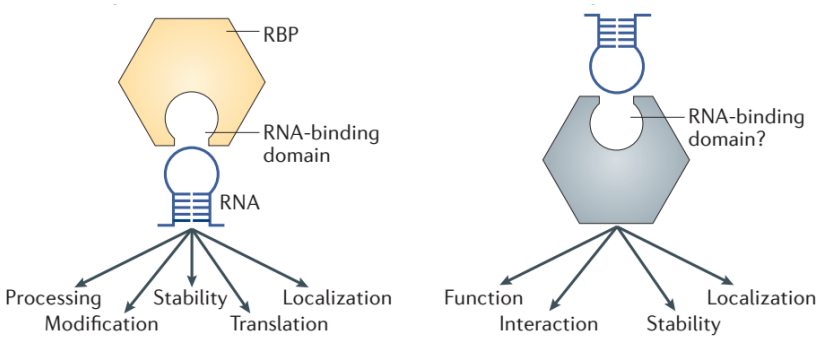

Figure 1.2: The functional relationship between RNA and RBPs. RBPs interact with RNAs by binding to specific *cis*-regulatory elements in a sequence or shape dependent manner, thereby modulating RNA fate. Inversely, interaction with RNA can also alternate RBP fate. Adopted from [45].

are differentially targeted by the respective RBP (**Fig. 1.3**). Recently, high-throughput SELEX (systematic evolution of ligands by exponential enrichment) studies with RBPs revealed that ∼70% of RBPs recognize short single stranded RNA sequences whereas ∼30% have a binding preference for structured RNA motifs. Interestingly, many RBPs can bind to multiple different motifs [48].

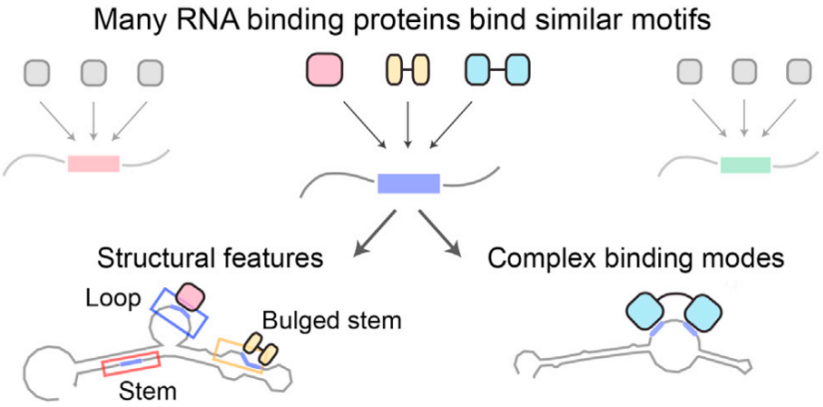

Figure 1.3: Composite RNA-binding domains and RBP binding sites guide RNA/RBP interaction. RBPs bind to RNA through a limited number of RBDs. Binding specificity is accomplished by RBD composition. Discrimination of similar RNA motifs is achieved through distinct affinities of an RBP to an RNA motif. The affinity is modulated by composite motifs, flanking sequence composition and RNA structure. Adopted and modified from [42].

The combination of canonical and non-canonical RBPs allows for a plethora of RBP-RNA interactions with a wide range of binding efficiencies and specificity, but not all of these can be assumed as relevant. A variety of RBP-RNA interactions is depicted in **Fig. 1.4**. One of the most common RNA-protein interactions in PTR is the high-

affinity binding of an RBP to a certain sequence or structure element in the target RNA. RBPs that recognize such signals often serve as adaptors for the recruitment effector proteins that modulate target mRNA stability or translation efficiency. In addition, numerous steps in RNA metabolism depend on indiscriminate RNA binding by RBPs. This is exemplified by various proteins involved in mRNA translation and degradation that need to bind non-selectively to RNAs to function. In a similar manner the exon-junction complex is non-selectively deposited at a fixed upstream position on the nascent mRNA. Approximately 50% of all RBP binding sites mapped in the RBDmap [49] map to intrinsically disordered domains (IDD) and ∼170 RBPs apparently rely exclusively on IDDs for RNA binding. This mode of binding is exemplified by the fragile X mental retardation protein 1 (FMRP). NMR studies revealed that glycine and arginine amino acid residues in the FMRP IDD guide the interaction of positively charged amino acids with Watson–Crick base pairs, which stack on two adjacent RNA G-quadruplexes that form as a result of protein–RNA co-folding. In addition to that, many shape-based

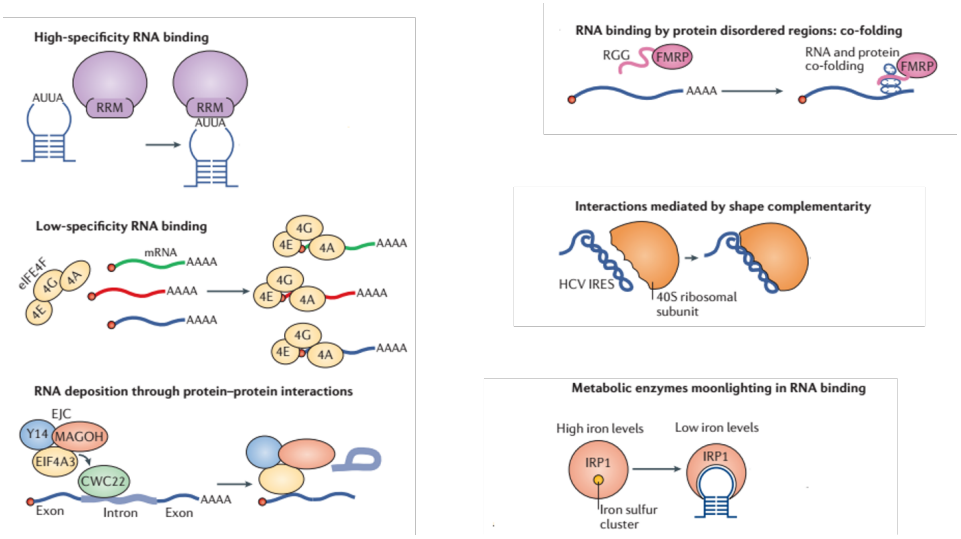

Figure 1.4: Different modes of RNA recognition by RBPs. RBPs harboring canonical RBDs interact with RNA in a sequence or shape dependent manner. EIF4E associates unspecifically with capped mRNAs in order to induce translation. The exon-junction complex is unspecifically deposited at exon junctions by interaction with splicing factor CWC22. The FRMP IDD cofolds with it target mRNA. The HCV IRES directly interacts with the 40 S ribosomal subunit by extended shape complementarity. The moonlighting RBP IRP1 binds with high affinity to conserved RNA stem-loop structures in iron-deprived cells. Adopted and modified from [45].

RNA-protein interactions are achieved through large RNA-protein interfaces without any defined protein regions. This is exemplified by the spliceosome, where snRNAs fold into complex structures that interact with protein partners that are complementary in structure and biochemical properties. Another example is the interaction of the

hepatitis C virus (HCV) internal ribosome entry site (IRES) with the ribosome, which again does not depend on a defined RNA-binding domain but rather on a strong shape complementarity between the IRES and the 40 S subunit of the ribosome. Over the last decade, mRNA interactome studies identified a number of metabolic enzymes with intrinsic RNA-binding activity. Iron regulatory protein (IRP) 1 (also aconitase, ACO1) requires an iron-sulfur cluster for its enzymatic activity, which however precludes RNA binding activity. In addition, to its role in the citric acid cycle, IRP1 regulates the expression of genes involved in the iron uptake and storage by binding to iron responsive elements in the UTRs of iron metabolism related genes (see section **1.2.1**)[45].

1.1.7 The Roquin Protein Family

The Roquin protein family contains the two RBPs, Roquin-1 and Roquin-2, which are key regulators of immune homeostasis. They repress the expression of immune relevant targets and thereby control gene expression and cell fate decisions. They are named by their unique RBD, the ROQ domain, that recognizes specific RNA stem-loop structures. Roquin localizes to P-bodies and stress-granules were it promotes target mRNA decay by the recruitment of the CCR4-CAF1-NOT deadenylase complex. Initially, Roquin-1 was identified by an N-ethyl-N-nitrosourea (ENU) mouse mutagenesis screen for regulators of autoimmunity. Homozygous mice with a point mutation in the Roquin-1 encoding *RC3H1* gene, resulting in a methionine 199 to arginine amino acid exchange (so-called *sanroque* mutation), developed a T follicular helper cell-driven lupus-like autoimmunity. This phenotype is induced by the activation of T cells and the resulting accumulation of T follicular helper (Tfh) cells and the production of autoantibodies [50] [51]. This is achieved by the induction of the Roquin target *ICOS* (inducible T-cell costimulator) in $RC3H1^{san/san}$ mice. Further immune relevant targets, like *CTLA4* and *Ox40*, cytokines *TNF* and *IL6*, transcription factors *IRF4* and *c-Rel*, and modulators of transcription *NFKBID* and *NFKBIZ*, have been subsequently identified as Roquin targets [52] [53] [54]. Roquin's role as a surpressor of immune relevant target genes is depicted in **Fig. 1.5A**. Interestingly, deletion of either Roquin-1 or Roquin-2 lead to perinatal lethality in mice. This suggests additional functions for Roquin proteins outside the immune response [53] [55].

Roquin-1 and -2 are highly similar and have ∼80% sequence identity in their N-terminal region. It was shown that Roquin proteins bind to the 3'UTR of their targets through their N-terminal half, encoding for a RING domain (really interesting new gene), which functions as an E3 ubiquitin ligase, and two RBDs: the ROQ domain and a CCCH ZnF domain. The effector functions of Roquin are harbored in the C-terminal region, which recruits the deadenylation machinery for subsequent target mRNA degradation [56] [52]. This part of the protein seems to be intrinsically disordered and contains a proline-rich stretch that might severe as surface for protein-protein interactions. In addition to this, a coiled-coil domain, which could mediate protein-protein interactions as well, has been predicted **Fig. 1.5B**.

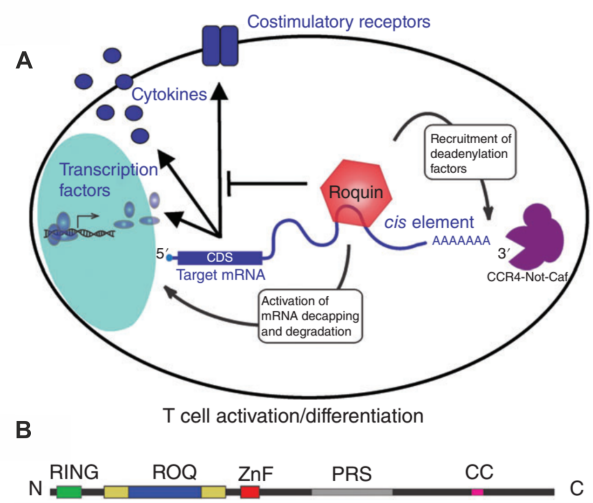

Figure 1.5: Roquins function in the immune response. (A) Roquin inhibits the expression of costimulatory T cell receptors, cytokines, and transcription factors. Roquin binds to *cis*-regulatory elements in the 3'UTR of target mRNAs and subsequently recruits the CCR4–CAF1–NOT deadenylation machinery. **(B)** Roquin-1 consist of an N-terminal RING domain, a ROQ domain and a ZnF. The C-terminal part of the protein contains a proline-rich sequence (PRS) and a short coiled-coil domain. Adopted from [57].

In contrast to many other RBPs that recognize single-stranded RNA motifs, Roquin proteins distinctively bind to an RNA motif that folds into a stem-loop structure, the so-called constitutive decay element (CDE) [56] [52]. Roquin's ROQ domain contains two RNA binding sites. The A-site specifically binds CDEs, whereas the B-site unspecifically recognizes dsRNA (**Fig. 1.6A**. The most extensively studied CDE is the initially discovered *TNF* CDE. This *cis*-regulatory element resides in the 3'UTR of the *TNF* mRNA and consist of a triloop hairpin structure. The structure of the *TNF* CDE bound by Roquin's ROQ domain is depicted in **Fig. 1.6B**). Detailed mutational analysis of the *TNF* CDE revealed Roquin's preference for a pyrimidine-purine-pyrimidine tri-nucleotide motif in the loop, along with two Y-P base pairs in the apical part of the stem, which also allows it to a purine stack of five consecutive purines at the 3' side of the stem [56]. However, this CDE consensus has been continuously revised as it does not match CDE-like hairpins found in other Roquin targets [58] [59]. A number of subsequent *in vitro* studies proposed that active CDEs form a 5-8 nt long stem capped by a U-rich tri-nucleotide loop [56] [60] [57]. The shape dependency of Roquin binding is further supported by structural analysis of Roquin-1 and Roquin-2 in complex with different CDEs. These *in vitro* studies propose a mainly shape-specific recognition of CDEs by Roquin that is independent of sequence [59] [61] [62] [63]. Moreover, SELEX experiments showed that Roquin binds to hairpins with U-rich hexaloops with high affinity. Such an alternative decay element (ADE) was identified in the 3'UTR of the Roquin target *Ox40* [64]. In addition to that, a recent publication suggests that Roquin might

bind linear sequence motifs as well [65].

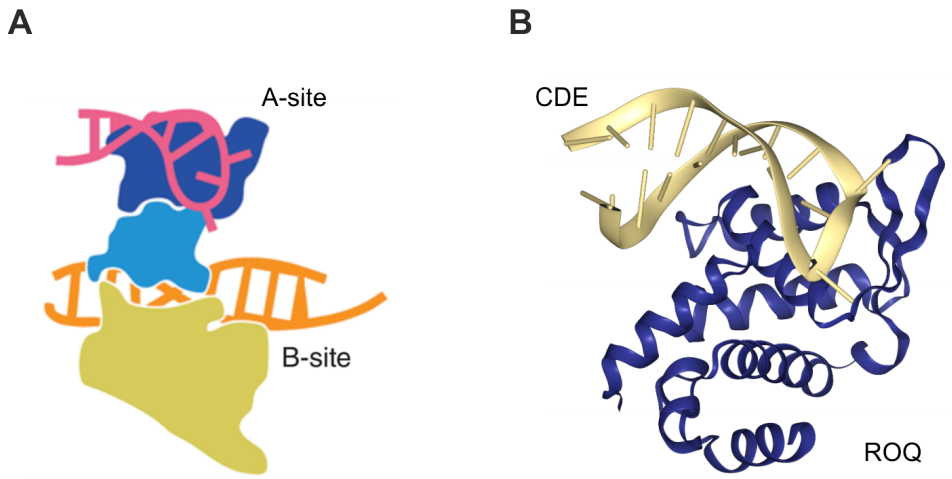

Figure 1.6: RNA recognition of Roquins ROQ domain. (A) Schematic representation of the ROQ domain RNA interaction. The A-site specifically recognizes CDEs, whereas the B-site unspecifically recognizes dsRNA.**(B)** Structure of the *TNF* CDE bound by the ROQ domain. Adopted and modified from [57].

To date it is still unclear what defines a potent *cis*-regulatory element for Roquin-mediated regulation. However, it seems that Roquin binding is much more versatile as the *in vitro* and *in vivo* studies on the *TNF* CDE suggested. Therefore, a much wider range of mRNAs than currently assumed could be functional Roquin targets.

1.2 The Role of UTRs in Gene Regulation

Post-transcriptional control of eukaryotic gene expression is an elaborate and extensive process. Essentially every step of the mRNA metabolism is regulated in an mRNA-dependent manner. The precise regulation of gene expression is a necessity for the correct temporal and spatial expression of genes. This regulation is primarily based on *cis*-regulatory elements residing in the UTRs of mRNAs, that are recognized by *trans*-acting factors in a sequence- or shape-dependent manner. This allows for complex PTR networks that enable cells to adapt to in- and extrinsic stimuli and thereby control the onset, duration or magnitude of specific beneficial or pathological activities. The structure of an generic mRNA encoding important *cis*-regulatory elements is depicted in **Fig. 1.7**. Altering of UTRs by alternative promoter usage or alternative polyadenylation can lead to serious pathologies [66]. The scientific field concerning the discovery of *cis*-regulatory RNA elements is divers and rapidly growing. The attempt to present all aspects of this field can only be incomplete. Therefore, I will concentrate on the core aspects that are involved in PTR which, in combination, make mRNA-based regulation an integral part of gene regulation.

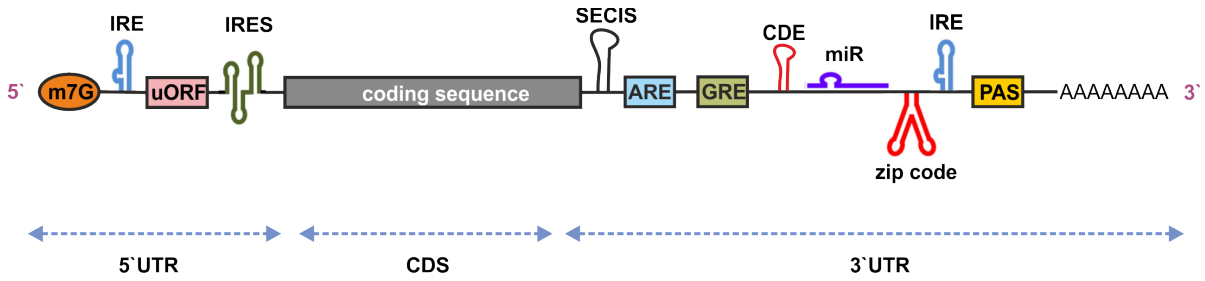

Figure 1.7: Structure of an eukarytic mRNA. Schematic representation of an mRNA molecule. The UTRs contain variable *cis*-regulatory elements that affect gene expression. m7G, 7-methyl-guanosine cap; IRE, iron responsive element; uORF, upstream open reading frame; IRES, internal ribosome entry site; CDS, coding sequence; SECIS, selenocystein insertion element; ARE, AU-rich element; GRE, GU-rich element; miR, microRNA binding site; zip code, RNA localization signal; CDE, constitutive decay element; PAS, polyadenylation signal. Adopted and modified from [67].

1.2.1 Regulation by 5'UTRs

The 5'UTR of a human mRNA has an average length of 100-200 nt and can harbor different functional elements that determine the translation of the downstream open reading frame. 5'UTRs of mRNAs that encode transcription factors, proto-oncoproteins, growth factors and proteins that are poorly translated under normal conditions, tend to be longer in vertebrates compared to other eukaryotes. An additional, conserved feature of vertebrate 5'UTRs is an increased GC-content of up to 60%, compared to 45% in 3'UTRs. The GC-content of RNA secondary structures can also affect translation efficiency independent of its thermodynamic properties [68] [69]. By means of alternative promoter usage, approximately 60% of all gene express alternative 5'UTRs and thereby contribute to transcriptome diversity [70].

Secondary structure

Secondary structures in 5'UTRs, such as stable stem-loop structures can be an important means of regulating translation. 90% of genes that encode for transcription factors, proto-oncogenes and growth factors have stable structured elements in their 5'UTR. 60% of these are located in close proximity to the 5' cap structure and are able to inhibit translation initiation. Stable hairpin structures near the 5' cap with a ΔG of -30 kcal/mol are sufficient to prevent binding of the pre-initiation complex to the mRNA. In addition, stable secondary structures can resist unwinding by the helicase EIF4A [69].

Upstream open reading frames

Upstream open reading frames (uORFs) are major regulators of gene expression. Computational predictions assume that approximately 50% of the human mRNAs contain

uORFs. uORF-mediated regulation of translation was verified for ∼100 transcripts, 30 of them in human. UORFs are characterized by an initiation codon in frame with a termination codon, located upstream or downstream of the main AUG. Consequentially translation initiates before the main ORF and a small peptide is formed. The 40S subunit of the ribosome can then either reinitiate and translate the encoded protein or dissociate without translating the main ORF. Thus, uORFs are generally associated with a reduction of protein expression. In addition to the translation inhibition of the main ORF, uORFs can also induce rapid mRNA degradation by the NMD mechanism as the uORF stop codon is recognized as premature. Interestingly, it was shown that in cellular stress response, the presence of uORFs favours the expression of certain mRNAs in the cell and thereby allows the respective mRNA to evade from global translation inhibition [71].

Internal ribosome entry sites

Internal ribosome entry sites (IRESs) occur in 5'UTRs of mRNA [71] and comprise a diverse group of *cis*-regulatory elements than can consist of a sequence of only 9 nt up to a strongly structured element of more than 600 nt in length [72][73]. Under stress conditions cap-dependent translation decreases. In order to maintain the translation of specific mRNAs under conditions of global translation inhibition, specific IRESs facilitate translation [74]. This process depends on RBPs that specifically interact with the IRES (ITAFs, IRES *trans*-activating factors) for the cap-independent recruitment of the ribosome. Initially, IRESs were identified in picornavirus RNA [75]. Since then, IRES elements were identified in a variety of human mRNAs encoding for proteins contributing to cell survival [76], proliferation [77], and angiogenesis [78].

Iron responsive elements

Iron responsive elements (IREs) comprise another important class or regulatory RNA elements that are located in the 5' and 3'UTRs of genes, important for cellular iron homeostasis. These ∼30 nt RNA stem-loop structures were initially discover in the 5'UTR of ferritin H- and L-chain mRNAs were they inhibit ferritin mRNA translation in iron-deprived cells [79]. Subsequently, IREs were discovered in the mRNAs of transferrin receptor, erythroid 5-aminolevulinic-acid synthase, aconitase, ferroportin, and divalent metal transporter 1. IREs are highly conserved for any given gene and are very similar between genes [80]. IREs interact with specific *trans*-acting RBPs, IRP1 and IRP2 with high specificity and thereby control translation efficiency and mRNA stability of the respective transcripts. If an IRE is located ≤40 nt downstream of the 5' cap, binding of the 43 S preinitiation complex is inhibited upon binding of IRP1 or IRP2. The 3'UTR of *TFRC* encodes five IREs that regulate mRNA stability. In iron-deprived cells IRPs bind to these stem-loops and thereby protect *TFRC* mRNA from endonucleolytic cleavage and subsequent degradation [80]. Binding activity of IRPs is directly regulated by the

iron concentration in the cytosol. If iron levels are high, IRP1 stores iron by binding to a [4Fe-4S] cluster, which also inhibits binding to IREs. IRP2 on the other hand is stabilized by the association with the [4Fe-4S] cluster under low iron concentrations [81]. Interestingly, IRP1 does not contain a canonical RBD and acts as a moonlighting RBP. In association with the [4Fe-4S] cluster it catalyzes an enzymatic reaction in the citric acid cycle.

1.2.2 Regulation by 3'UTRs

3'UTRs control protein levels by altering mRNA stability and translation efficiency. This regulation is most commonly achieved by *cis*-elements. In addition, 3'UTRs can facilitate local translation by the regulation of mRNA localization. mRNA fate can further be modulated by the specific inclusion or exclusion of *cis*-regulatory elements in 3'UTRs by alternative polyadenylation as well as the observation that 3'UTR cleavage products can act as ncRNA.

In human genes the median 3'UTR length is approximately 1200 nt. Although 3'UTRs are less conserved than the protein-coding regions of an mRNA, they still show a high degree of sequence conservation across vertebrates. Similar to the protein-coding region 3'UTRs contain island of sequence conservation, and these island often represent RBP and miR binding sites [46]. In humans, 51–79% of genes express alternative 3'UTRs. Alternative 3'UTRs can increase the accessibility of regulatory elements, that would otherwise be structurally constrained within a long UTR. An important aspect of the mRNA life cycle is the correct mRNA localization which is usually mediated by localization signals (zip codes) in 3'UTRs. These elements are recognized by RBPs that interact with motor proteins that move mRNAs along the filament of the cytoskeleton. This is exemplified by the *ASH1* mRNA which contains four *cis*-elements including one located in the 3'UTR, important for proper localization. These elements are bound by RBPs that recruit a myosin motor protein, which facilitates the movement along actin microfilaments. Similar mechanisms have been described for the majority of localized mRNAs [46].

1.2.3 AU-rich Elements

AREs are *cis*-regulatory elements that modulate mRNA stability and translation and are predominantly encoded in the 3'UTRs of early-response genes and transiently expressed mRNAs [82]. These mRNAs typically encode proteins that regulate cell proliferation, RNA metabolism, transcriptional regulation, signaling, and response to stress and microbes. Such genes require a very precise control to be able to instantly react to such stimuli. Therefore, in addition to transcriptional control, regulation of translation and stability of the mRNA is a necessity. The importance of ARE-mediated repression is highlighted by the observation that aberrant expression of ARE-containing mRNAs is associated with a plethora of pathological states, i.e. cancer, chronic inflammations and

autoimmune pathologies [83]. Initially, AREs were identified in the 3'UTRs of unstable mRNAs encoding cytokines and lymphokines [84]. To date, >4000 ARE-containing mRNAs have been identified in the human genome, but only few of them have been experimentally verified [85] [86]. This makes AREs the most common *cis*-regulatory element. The canonical sites for AREs are 3'UTRs, but recent studies showed they are also abundant in introns were their sequence is much more complex compared to 3'UTR AREs [85]. AREs are classified into distinct families, which presumably are recognized by specific sets of RBPs that act in *trans* for differential regulation. AREs are 50-150 nt long and contain the core -AUUUA- motif embedded in an AU-rich sequence context [85]. Based on the occurrence of the pentamer motif and overall sequence, AREs are devided into three classes. Class I AREs contain several scattered copies of the -AUUUA- pentamer in a U-rich environment. Class II AREs contain at least two overlapping UUAUUUA(U/A)(U/A) nonamers. Class III AREs are less well defined U-rich elements without the core pentamer motif. An example for a class III ARE can be found in the *c-jun* mRNA [83].

AREs are recognized by a variety of RBPs, so called ARE-BPs. Among these HuR, a protein from the embryonic lethal abnormal vision-like protein family, is a ubiquitously expressed ARE-BP. It is a universal RNA stability factor. It is known to bind to AREs in 3'UTRs and introns, and to stabilize transcripts and enhance translation. Another group of important ARE-BPs are proteins from the CCCH ZFP36 family, which contains tristetraprolin (TTP). Because of its strong anti-inflammatory role and its role as tumor suppressor, it is the most widely studied ARE-BP. TTP only binds to class II AREs and promotes mRNA decay. The first ARE-BP that was discovered, AU-rich-binding factor 1 (AUF1, also hnRNP D) and its isoforms, bind to AREs and modulate mRNA stability. In addition, AUF1 can interact with HuR and TTP. T cell intracellular antigen-1 (TIA-1) and TIA-related (TIAR) proteins cause translational inhibition by binding to AREs.

Specific binding of an RBP to an ARE is determined by the type of ARE, its length and context. Specificity is further regulated by ARE-BP modifications and interaction between ARE-BPs. In this context competitive binding, as in the case of TTP and HuR, and cooperative binding, e.g. TTP and KSRP, significantly contributes to target mRNA fate [85].

In addition to AREs, other classes of U-rich regulatory elements that are associated with rapid mRNA decay, exist. Computational analyses estimate that 5% of human mRNA contain GU-rich elements in their 3'UTRs [87]. These elements are highly conserved among mammals. These *cis*-regulatory elements contain 2-5 overlapping -GUUUG- pentamer motifs. Members of the CELF protein family have been shown to recognize GU-rich elements. GU-rich-elements are functionally linked to mRNA decay, polyadenylation and pre-mRNA translation [88] [89].

1.2.4 Structured Regulatory RNA Elements

A plethora of single stranded RNA elements is known and accessible in a variety of databases. There are databases for annotated binding sites of RBPs [90], miRs [91] as well as AREs [92] and splicing regulators [93]. In contrast, only a small number of structured regulatory RNA elements, that regulate mRNA stability or translation efficiency, is known. In addition to the few individual examples that exist, effort is put into the computational prediction of structured regulatory RNA elements [94].

RNA can adopt abstract shapes by folding into complex secondary structures by intricate patterns of Watson-Crick or G-U base pairing. A number of non-canonical base pairing events involving Hoogsteen base pairing and sugar-edge patterns and the formation of triple base pairs contribute to the formation of extensive tertiary structures. Common structural features of RNA molecules include RNA-RNA duplexes, single-stranded regions, hairpins, bulges, internal loops and junctions. In addition, inter-base pairing, such as loop-loop interactions and pseudoknots can form tertiary structures [95].

A important example for a structured RNA element that regulates mRNA stability is the histone RNA stem-loop [96] [97]. In constrast to poly(A)-mRNA, histone mRNAs contain a conserved 25-26 nt stem–loop structure at their 3' end [98]. This structured RNA feature is recognized by the stem-loop binding protein (SLBP), which induces rapid mRNA decay of histone mRNAs at the end of S-phase or when DNA synthesis is inhibited. Upon binding, SLBP recruits additional proteins, which faciliate the addition of an oligo(U) tail to histone mRNA. This initiates the decapping of the histone RNA and the subsequent 5'→3' degradation, whereas the 3'→5' degradation via the exosome stalls [98]. Recent studies revealed, that exoRNase Eri1 binds to the histone stem-loop and trims the mature histone mRNAs 3'ends. Subsequently the oligo(U)-tail is recognized by Lsm1-7 which then interacts with Eri1, which is then able to degrade the stem-loop [99].

The RBP Smaug recognizes RNA stem-loop elements in *Drosophila* maternal mRNAs. Binding of Smaug induces the CCR4-CAF1-NOT-dependent mRNA degradation and also regulates the translation of *nanos* mRNA in flies [100].

Another example for an RNA structure that targets mRNAs for degradation is the Staufen binding site (SBS). Staufen-1 targets mRNA for degradation by binding to long RNA-RNA duplex [101] in a process call Staufen-mediated decay (SMD). Therefore, Staufen-1 interacts with the NMD factor UPF1 to elicit mRNA decay. ∼1% of HeLa mRNAs are targets of Staufen-1 suggesting that SMD constitutes a significant post-transcriptional regulatory pathway [102]. Staufen binds to its targets through four dsRBDs. SMD requires an SBS that is situated sufficiently downstream of a termination codon, in the 3'UTR. SBSs are formed by intramolecular base pairing within an mRNA 3'UTR or intermolecular base pairing between an Alu element and a partially complementary Alu element within certain lncRNAs [102] which allows transactivation

of SMD by lncRNAs [103].

Selenium is considered an essential trace element in human health and is incorporate in ∼25 human selenoproteins via the 21st essential amino acid, selenocysteine. This depends on the translational recoding of an in-frame UGA codon. 3'UTRs of selenoprotein mRNAs always harbor a selenocysteine insertion sequence (SECIS) downstream of the UGA codon. SECIS elements are not conserved at sequence level, but their secondary structures can be represented as a consensus stem-loop motif. SECIS elements are recognized by SECIS binding protein 2 (SBP2) and ribosomal protein L30, but little is known about the exact mechanisms of UGA recoding [104].

In addition to this CDEs and IREs are conserved RNA hairpin structures that destabilize their target mRNAs or inhibit their translation and have been described in sections **1.2.1** and **1.1.6**.

1.2.5 Prediction of Regulatory RNA Elements

To understand how certain expression patterns emerge, it is necessary to know the exact number and composition of regulatory elements in a particular mRNA. This knowledge is also the key to understanding how these expression patterns change in response to a variety of stimuli. A number of methods are available for the prediction and high-throughput identification of linear sequence motifs [105] [106] [107]. As basically all functions of RNA are controlled by its ability to fold into complex secondary and tertiary structures the limited explanatory power of these methods can be easily appreciated. In contrast to linear sequence motifs, the identification of structured RNA elements that function in gene regulation is challenging and therefore less studied. To overcome this drawback, methods for the *in vivo* probing of global RNA structure have been developed recently. These methods are based on cell permeable chemicals that covalently modify the RNA with different reactivity for single- and double-stranded regions and lead to the displacement of the reverse transcriptase (RT) during cDNA synthesis. These RT-stop sites are than analyzed by high-throughput sequencing and yield a high resolution map of RNA structure based on the relative reactivity of the modifying chemical towards individual nucleotides [108] [109] [110] [111]. However, these averaged reactivity profiles fail to capture the complexity of RNA structures that include long-range structures, pseudoknots, and alternative conformations and no assumptions can be made whether regions with low reactivity are actually double-stranded or not accessible for the chemical because the RNA is bound by a protein. Recent approaches circumvent this caveat by the direct detection of double-stranded regions by proximity ligation after cross-linking. In these methods, unbound double-stranded regions are detected by psoralen-induced cross-linking [112]. In contrast, protein-bound regions can be detected by UV-induced cross-linking in combination with immunoprecipitation of the desired protein [101]. While these methods are promising tools for the comprehensive elucidation of the *in vivo* genome-wide structural landscape of RNA, they have some inherent limitations. The major disadvantage of all these methods is that they only provide a

snapshot of RNA structures present in the moment of the experiment. However, many RNAs with crucial roles in cell fate decisions are only expressed in certain cellular states. Therefore, cell-type specific, cell-cycle specific or stress-induced RNAs are usually not detected with the above mentioned methods [113].

An alternative approach to detect specific RNA structures, that is independent of the expression level of RNAs, is the computational prediction of RNA secondary structures [114] [115]. By means of RNA structure prediction it is possible to query simultaneously the whole genome or certain features thereof, like UTRs, for potential RNA secondary structure formation. In addition, base pairing probabilities can highlight base pairs and structures that are more likely to be formed compared to base pairs with lower probability of pairing [116] [117].

Conserved elements in the UTRs generally serve as an indicator of functional importance. This can be conserved binding sites for miRs and RBPs or conserved structured RNA elements, like IREs. Therefore, the evolutionary conservation of structure is equally indicative of functional importance [118]. Assessing structural conservation in UTRs therefore improves the *de novo* discovery of RNA structures that function in gene regulation.

Nucleotide mutation patterns that do not disrupt RNA structure formation are called covariations. A structurally conserved RNA tends to show a pattern of compensatory mutations consistent with some base-paired secondary structures [119]. Compensatory evolution, as mediated by RNA secondary structure, can result in detectable patterns of nucleotide substitutions in the phylogenetic alignment of homologous RNA sequences and is indicative for functional structures. Based on this phylogenetic-comparative method, consensus structures for tRNAs, rRNAs, and ribozymes have been generated[120]. This principal was first used by the program QRNA. QRNA uses mutational patterns in phylogenetic alignments to identify conserved structures [119]. Based on covariation, computational identification of functional, structured RNA elements in bacterial mRNA yielded >100 candidate structured RNAs [121]. The large number and variety of available bacterial genomes together with their lower sequence content is a decisive advantage for the utilization of covariation to identify conserved RNA structures. In contrast, covariation is less useful for the *de novo* discovery of conserved RNA structures in mammals, because of the smaller number of available genomes, which are less diverse as mammals have a closer evolutionary relationship. An alternative method for the *de novo* discovery of conserved RNA structures is folding stability. This principal was initially used by RNAz to find conserved RNA structures by combining a measure for thermodynamic stability with a measure for structure conservation. In contrast to previous methods, this approach is independent from mutational patterns [122] [123]. One drawback of RNAz is that it relies on a fixed input alignment and therefore favors sequences with high nucleotide identity over sequences that are difficult to align. To solve this problem, the programs Dynalign and Foldalign were extended to discover conserved structures [124] [125]. Both programs are based on the Sankoff algorithm

and use dynamic programming. They take two unaligned sequences as input which are simultaneously folded and aligned to identify the lowest free energy conserved structure [126] [127].

1.3 Aim of this Study

The aim of this thesis is the identification and characterization of novel, structurally conserved RNA elements in the 3'UTRs of human genes. Compared to single stranded RNA motifs that are involved in PTR, only few examples exist for structurally conserved RNA motifs. Structured *cis*-regulatory RNA elements can specifically regulate complex genetic circuits and thereby ensure cellular homeostasis in response to a variety of stimuli [56] [80] [96]. By the *de novo* identification of structurally conserved RNA elements, I want to expand the current knowledge of how extensively this important type of *cis*-regulatory elements contribute to the regulation of gene expression.

Therfore, I analyzed computationally predicted, structurally conserved RNA elements in 3'UTRs. I investigated the regulatory properties of predicted elements and performed structure function analysis to identify crucial RNA moieties in one regulatory element from the 3'UTR of the *UCP3* mRNA. For the identification of *trans*-acting factors, I enriched RNA specific interacting proteins and verified Roquin as regulator of *UCP3* mRNA levels. The identified CDE-like elements were then used to predict and verify new targets of Roquin that strongly suggest additional functions for this important regulator outside the immune response.

2 Results

The central hubs of RNA regulation are mRNAs. Basically every step of the complex life of an mRNA is post-transcriptionally regulated. This regulation is based on *cis*-regulatory elements located in the UTRs of the mRNA. These elements in turn are recognized by so called *trans*-acting factors like non-coding RNAs, i.e. miRNAs, and RNA binding proteins in an either sequence- or shape-dependent manner. In particular, regulation of mRNA stability is an important means of a cell to quickly respond to certain extrinsic or intrinsic stimuli and stress response. Prominent *cis*-regulatory elements in mRNA stability regulation include linear sequence motifs like miR binding sites which, by base pairing, lead to RISC assembly and subsequent RNA degradation [29] and AU-rich elements, which are recognized by ARE-binding proteins that destabilize the RNA by recruiting the CCR4-NOT deadenylase complex[46]. In contrast, small RNA hairpin structures, like SBSs, CDEs and IREs are recognized by RNA-binding proteins STAU-1, RCH3H1/2 or aconitase in a shaped-dependent manner.

Roquin binding elements, so-called CDEs, have been identified in the 3'UTRs of *TNF*, *ICOS*, *OX40* [64], *HMGXB3*, *IL6*, *IER3*, *NFKBID*, *NFKBIZ*, *PPP1R10* and *TNFRSF4* [62], with only the *TNF* CDE being extensively characterized so far [56]. In my doctoral thesis I describe the characterization of a tandem CDE structure in the 3'UTR of the mitochondrial anion carrier *UCP3*, which is associated with fatty acid metabolism, obesity and insulin resistance [128][129]. I discovered the repressive element by a computational analysis for structurally conserved elements in 3'UTRs. After in depth functional and structural analysis of the element, I identified Roquin as the *trans*-acting factor. I could show that both CDEs are targeted individually by Roquin, but the presence of both CDEs is necessary for robust regulation. Both CDEs residing within the element deviate from the *TNF*-derived consensus. Detailed mutational analysis *in vivo* allowed me to redefine the binding preferences of Roquin. Using this data allowed me to accurately predict a set of new high potential Roquin targets genes and confirm new targets *in vivo*. Taken together, this allowed me to expand the established role of Roquin in immune response and inflammation to additional, yet unstudied cellular functions of Roquin.

2.1 Identification of Regulatory RNA Elements Based on Structural Conservation

Structured RNA elements have been known to be involved in many RNA regulation processes like pre-mRNA splicing [130], regulation of miR target site accessibility [131] and mRNA stability [102]. In order to understand how complex expression patterns are achieved, knowledge about type and number of regulatory elements in a UTR is necessary. Yet, so far, only a small number of regulatory mRNA structures has been identified compared to linear RPB binding sites or miR target sites. This can be explained by the fact that it is more difficult to identify structured RNA elements than linear RNA elements. In this part of my thesis, I show how structurally conserved RNA elements can be identified with the algorithm Dynalign. Using Dynalign, I identified a repressive element in the 3'UTR of the *UCP3* mRNA. In depth functional and structural analysis revealed a 64 nt minimal motif that folds into a tandem RNA hairpin structure that is sufficient and necessary for the regulation of gene expression.

2.1.1 A Conserved Structure in the 3'UTR of *UCP3* Reduces Gene Expression

The initial purpose of my study was the *de novo* identification of regulatory RNA structures. In collaboration with the Mathews Lab (Rochester University, NY), we investigated the conservation of RNA structures on a genome-wide scale. Therefore, we compared all UTRs of human and mouse. We selected the mouse genome for this comparison, because it is currently the best annotated genome besides the human genome. In addition, the goal of our analysis was to exclusively rely on RNA secondary structure conservation. Therefore, we performed a bioinformatic prediction of conserved RNA structures using the RNA folding program Dynalign. Dynalign simultaneously folds and aligns two input sequences and thereby significantly improves the accuracy of the structural predictions [132] [133]. It predicts the conserved secondary structure with the lowest total score, which is the sum of the folding free energies of the two structures as estimated by nearest neighbor rules and a score for the alignment [113]. Importantly, this alignment is made to reflect the conserved structure, and does not consider sequence identity. Therfore, the accuracy of structure prediction is not adversely affected by sequences with low similarity [133] [125]. This is a strong advantage over other RNA structure prediction programs that rely on fixed sequence alignments as input. Accurate prediction by such programs very much depends on the quality of the initial sequence alignment and therefore favors sequences with higher similarity [134]. In contrast, Dynalign includes the advantage of evolutionary conservation for structure prediction, without the necessity for high sequence conservation. For the discovery of structured RNA elements in genomes, machine learning was used to train a model to classify Dynalign output [125] [135]. Two sequences are input to Dynalign, the best

scoring common structure is predicted and features of the sequences (length and nucleotide composition) and the estimated folding free energy change are used to estimate the probability that the two sequences represent a conserved and stable RNA secondary structure.

For the genome-wide search, all human UTRs were divided into 100 nt long windows, with a step size of 50 nt and structurally compared to the corresponding mouse sequences. 100 nt long windows were chosen, to balance structure prediction accuracy and scanning efficiency. Using this approach, ∼5900 genes were predicted to contain at least one structurally conserved element in their 5' or 3' UTR, with >0.9 estimated probability of being a conserved structure. Importantly, known iron responsive elements in the 3'UTR of the transferrin receptor (TFRC) have been correctly predicted. In order to test the regulatory function of the predicted windows, the 100 nt long sequences were fused to the 3' end of a firefly luciferase. To quantify the influence on luciferase activity dual luciferase assays were performed and the relative luciferase activity of the firefly luciferase fusions compared to an empty vector control without a predicted window.

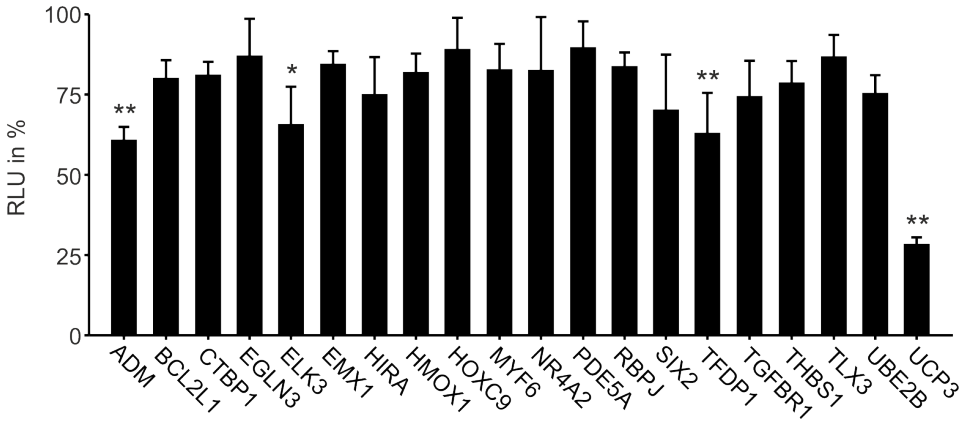

Figure 2.1: Structurally conserved regions predicted by Dynalign influence gene expression. Luciferase activity of various 3'UTR fusion constructs. 100 nt long 3'UTR sequences predicted to encode structurally conserved regions, were fused to firefly luciferase. Firefly luciferase activity was normalized to *Renilla* luciferase as internal transfection control. Values are normalized to an empty vector control. n = 3. (**) P-value <0.01. (*) P-value < 0.05. Adopted and modified from [113].

For an initial testing, I selected 20 windows for regulation of gene expression based on an estimated folding probability of >90% to be a conserved structure. Also the total length of the UTR containing a predicted window should fall within the average UTR length of 500 to 1500 nt of human 3'UTRs. In order to simplify functional analysis the predicted windows were not allow to overlap with other windows. The selected 100 nt long 3'UTR sequences were cloned downstream of the firefly luciferase reporter gene into the pDLP plasmid. The pDLP plasmid also encodes a *Renilla* luciferase which is expressed from a CMV promoter under control of the same enhancer element as the

firefly luciferase and was used for internal normalization. The pDLP plasmid map is depicted in **Suppl. Fig.5.1**. To investigate the influence of the elements on luciferase activity, luciferase-3'UTR fusions were transiently transfected into HEK293 cells and luciferase activity measured (**Fig. 2.1**).

Four of the 20 windows significantly reduced luciferase activity in HEK293 cells. The predicted structurally conserved windows from the 3'UTRs of *ADM*, *ELK3* and *TFDP1* reduced luciferase activity to $\sim 60\%$ compared to the empty vector control in HEK293 cells. The most efficient reduction was observed with the window from the 3'UTR of *UCP3* which reduced luciferase activity to $\sim 35\%$ in HEK293 cells. The *UCP3* gene codes for a mitochondrial membrane protein. The UCP3 protein is implicated in fatty acid metabolism, obesity and insulin resistance (reviewed in [128] [129]), yet its physiological function is still under debate [136] [137]. Interestingly elevated UCP3 levels were observed in obesity-resistant mice and a decrease in UCP3 levels was associated with insulin sensitivity, a condition that precedes diabetes. Thus, modulation of UCP3 levels is an interesting target for the development of novel therapeutics.

To further confirm the importance of this element in *UCP3* gene expression the complete 1131 nt long *UCP3* 3'UTR as well as a deletion mutant without the 100 nt long window was cloned into the pDLP plasmid. In addition, the influence of the 100 nt long element alone, duplication and randomization of the element were investigated. An overview of the constructs is depicted in **Fig. 2.2A**. To investigate the influence of the

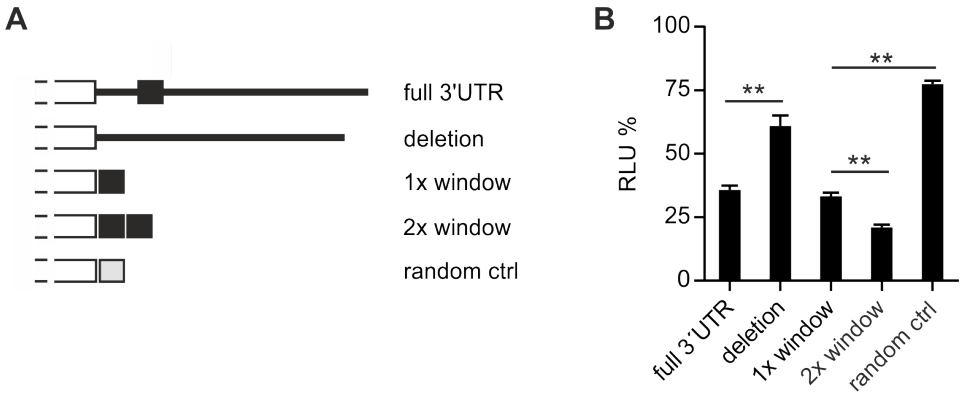

Figure 2.2: The 100 nt long, structurally conserved region in the 3'UTR of *UCP3* codes for a repressive element. (A) Overview of luciferase reporter constructs. Different fragments of the *UCP3* 3'UTR were fused to the firefly luciferase. **(B)** Luciferase activity of UCP3 3'UTR fusion constructs shown in (A). Firefly luciferase activity was normalized to *Renilla* luciferase as internal transfection control. n = 3. (**) P-value < 0.01. Adopted and modified from [113].

various 3'UTR constructs on luciferase activity, HEK293 cells were transiently transfected with the corresponding pDLP reporter constructs. 24 h post transfection firefly luciferase activity was measured. The results of the reporter gene assay are depicted in **Fig. 2.2B**. The complete 3'UTR and the 100 nt window alone repress luciferase activity

in HEK293 cells to the same extend. Duplication of the regulatory window further emphasizes the repressive effect on luciferase activity. Deletion of the 100 nt long window significantly restores luciferase activity, as does randomization of the 100 nt window. In conclusion this showed that the 100 nt long structurally conserved region predicted by Dynalign is necessary and sufficient for effective post-transcriptional regulation by the 3'UTR3 of *UCP3*.

2.1.2 The reduction of gene expression is not induced by miR-152

MiRs are a class of short, ∼22 nt long, endogenous non-coding RNAs that post-transcriptionally control gene expression. In addition to RBPs, they represent the most important group of regulatory molecules that have a decisive influence on gene expression. In higher eukaryotes it is predicted that miRs control ∼ 60% of all protein-coding genes and have been shown to participate in every cellular process investigated so far [36]. MiRs repress their target genes by either target mRNA destabilization or by repression of their translation [39]. To exclude the possibility that the repressive effect was induced by miRs I evaluated the structurally conserved region with the TargetScan algorithm [138]. This algorithm predicts miR binding sites in a given sequence. The computational analysis revealed two overlapping conserved target sites for the miR families miR-148a/b and 152 and miR-130a/b and 301a/b (**Fig.2.3A and B**).

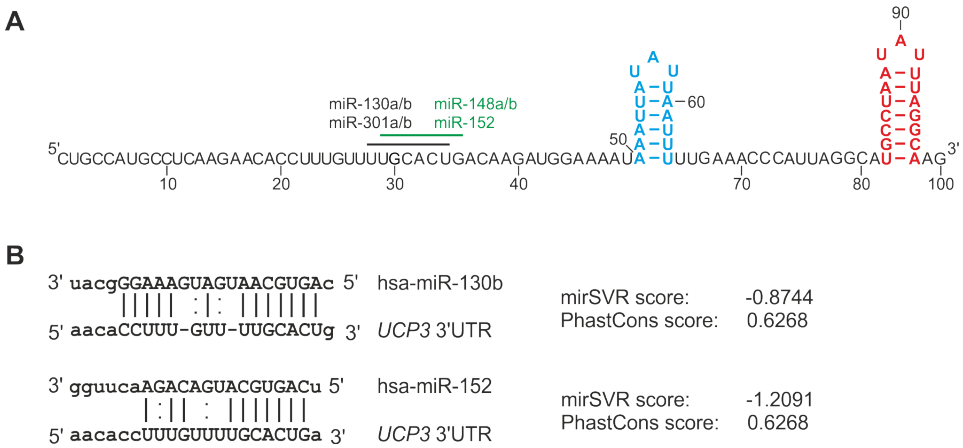

Figure 2.3: The structurally conserved region of the UCP3 3'UTR encodes a functional miR binding site. (A) Location of the miR binding site within the Dynalign window. The window encodes two overlapping sites indicated by black and green lines, respectively. **(B)** Predicted miR target sites by *miRanda* [139]. Shown are the target site alignment as well as *miRanda-SVR* calculated miR-SVR scores and target site conservation by *PhastCons*. Adopted and modified from [113].

To assess whether these miRs are expressed in HEK293 and HeLa cells, I performed a quantitative stem-loop RT-qPCR. In this assay, a miR specific stem-loop oligonucleotide

was hybridized to the miRNA of interest. The hybrid was then reverse transcribed to cDNA [140]. In the next step, the product of reverse transcription was then quantified by qPCR. The values were normalized to the small nucleolar RNA U48 as internal control. All of these seven miRs were either undetectable or expressed at very low levels in HEK293 and HeLa cells (**Suppl. Tab. 5.3**). Therefore, it was highly unlikely that they were responsible for the observed reduction in luciferase activity. To further proof that the repression by the *UCP3* element is not triggered by these miRs, a luciferase construct was cloned in which the miR binding site was deleted. Deletion of the target site had no effect on luciferase activity (**Fig. 2.4 A**).

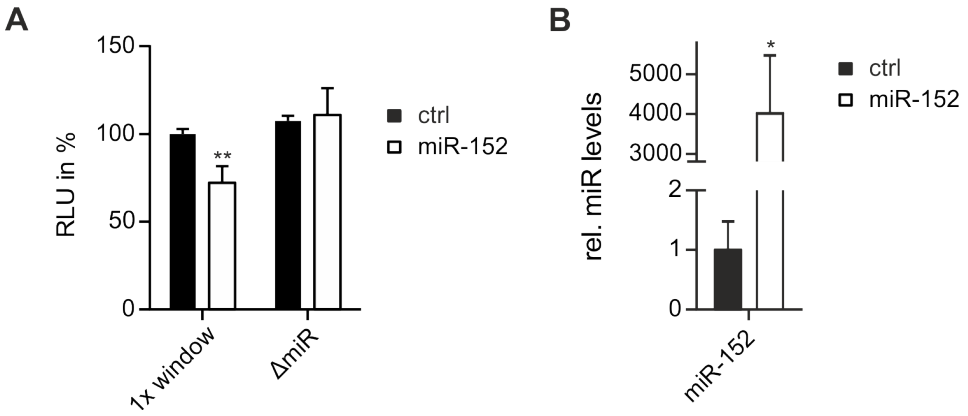

Figure 2.4: Influence if miR-152 on the repression by the *UCP3* element. (A) Luciferase activity of the *UCP3* Dynalign element (1x window) and a construct in which both miR binding sites are deleted (ΔmiR) with and without overexpression of miR-152. Firefly luciferase activity was normalized to *Renilla* luciferase as internal transfection control. Values are normalized to an empty vector control. n = 3. **(B)** RT-qPCR quantification of miR-152 levels after overexpression. Values are normalized to the snoRNA U48 as internal control. (**) P-value > 0.01. (*) P-value > 0.05. Adopted and modified from [113].

To investigate whether the predicted miR binding sites are functional at all, I investigated whether overexpression of miR-152 has an influence on the *UCP3* reporter construct. Of all predicted miRs that can bind to the target site miR-152 has the highest miRSVR score, which is a measure for the extent of their expected downregulation at the mRNA or protein level [139]. Also miR-152 was chosen because of its association with type 2 *diabetes mellitus* [141] and its ability to alter dynamics of the intracellular ATP/ADP ratio [142], which is a principal function of UCP3. For miR-152 overexpression experiments, the genomic locus of miR-152 together with its endogenous sequence context (200 nt up- and downstream of the mature miR) was cloned into the plasmid pCMV-MS and expressed from a CMV promoter. To investigate the influence of miR-152 on luciferase activity, HEK293 cells were cotransfected with either the complete *UCP3* window or the miR target site deleted window containing pDLP plasmid and

pCMV-miR152. 24 h post transfection, luciferase activity was measured (**Fig.2.4A**). Overexpression of miR-152 was verified by stem-loop RT-qPCR (**Fig.2.4B**). Overexpression of miR-152 reduced luciferase activity of the *UCP3* wild type (wt) sequence significantly, but not that of the deletion construct without miR target sites. From those findings I concluded that the observed effect was not caused by miRs in HEK293 cells. However, UCP3 may be regulated by miRs in other cell types or under certain stress conditions as the overexpression experiments showed that, in principal, miR-152 can repress gene expression of *UCP3*.

2.1.3 A 64 nt Minimal Motif is Sufficient for Gene Repression

Since I could now exclude that the repression of luciferase activity is caused by miRs, I wanted to identify the features of the *UCP3* window that are important for regulation. *Dynalign* predicted a structure that consists of three RNA hairpin structures that are separated by regions that are less conserved at both, structure and sequence level (**Fig. 2.5**). To asses the importance of the three hairpin structures in regulation of gene ex-

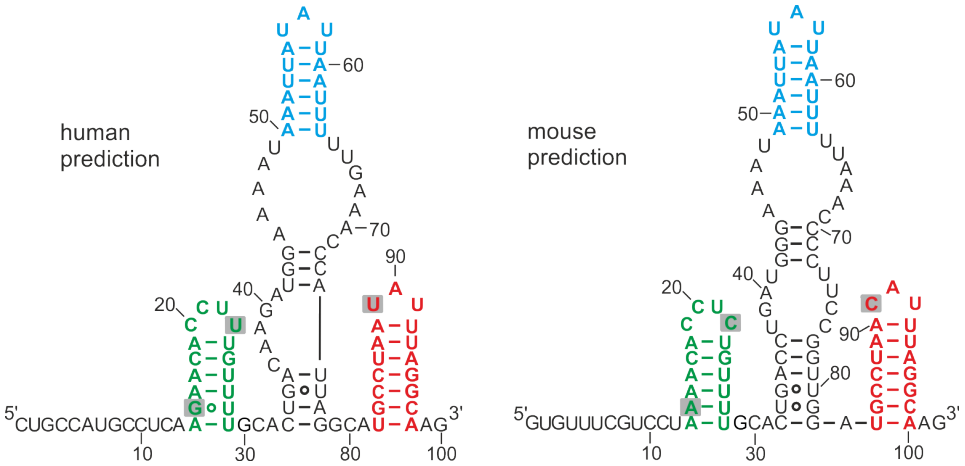

Figure 2.5: Secondary structure prediction by Dynalign. Secondary structure prediction by Dynalign of the structurally conserved regions in the human and mouse *UCP3* 3'UTRs. Geen: Hairpin 1, Blue: Hairpin 2; Red: Hairpin 3. Deviating nucleotides in the three hairpins are highlighted in grey. Adopted and modified from [113].

pression, truncation studies of the 100 nt long *UCP3* window were performed. Therefore, the element was shortened step by step from the 5' or 3' end and the luciferase activity was measured. An overview of the truncations is depiced in (**Fig. 2.6A**). Deletion of hairpin 1 alone or together with the miR target sites had no effect on repression. However, truncation or deletion of hairpin 2 or 3 led to a significant increase in luciferase activity to $\sim$ 70% (**Fig.2.6B**). The shortest version that still has the full regulatory function (del2) consists of the 64 nt at the 3' end of the *UCP3* window and contains RNA hairpins 1 and 3. This minimal motif is very AU-rich (73.4%) and harbors the

core pentamer motif -AUUUA- typically found in AREs. AREs are important regulators of gene expression. It has been shown that RBPs like HuR, TTP, TIA-1, KSRP and AUF1 can bind to such elements and reduce gene expression either by reducing mRNA half-life or translation efficiency [83]. To investigate this further a reporter construct was cloned in which the -AUUUA- motif was mutated to -AGGGA-. Mutation of the pentameter led to a significant increase in luciferase activity (see MUTII in **Fig. 2.10B**. However, the mutation not only altered the sequence but also the secondary structure of hairpin 3. Therefore, the effect could not be clearly attributed to the pentamer motif. Furthermore, the mutation does not explain why the deletion of hairpin 2 also leads to a loss of function. To exclude any unspecific luciferase reduction by the high AU-content, we fused five different randomized sequences with the same nucleotide composition and one randomized sequence with uniform nucleotide content (64 nt) to the luciferase reporter. All randomizations significantly restored luciferase activity (**Suppl. Fig. 5.7**). This confirmed, that the 64 nt long del2 is the minimal motif that is necessary and sufficient for regulation.

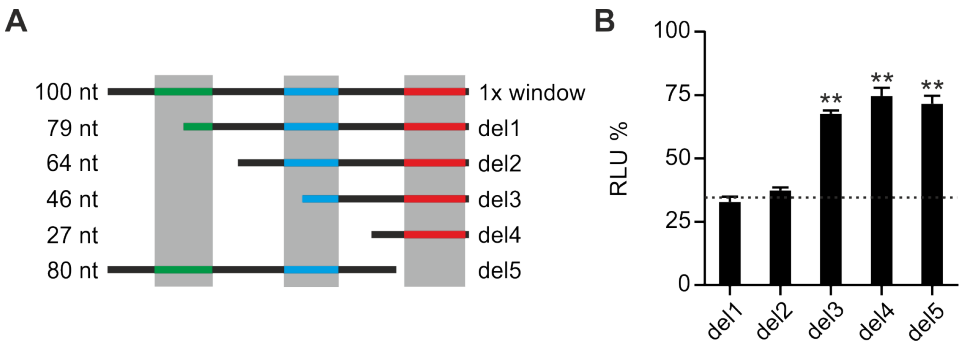

Figure 2.6: Truncation study of the 100 nt *UCP3* window. **(A)** Overview of truncations of the 100 nt long window. **(B)** Luciferase activity of *UCP3* truncation constructs shown in (A). Firefly luciferase activity was normalized to *Renilla* luciferase as internal transfection control. Values are normalized to an empty vector control, without *UCP3* UTR sequences. Luciferase activity of the 100 nt window is indicated as dashed line. n = 3. (**) P-value <0.01. Adopted and modified from [113].

2.1.4 The Repressive Element is a Conserved Tandem Stem-Loop Structure

In the next step, it should be investigated whether the observed effect is caused by a linear sequence motif or by the RNA structure predicted by Dynalign. To address the structure experimentally, I performed in-line probing analysis. In this RNA cleavage assay, base-paired nucleotides are protected from spontaneous phosphodioester bond hydrolysis. Radiolabeled *UCP3* del2 RNA was incubated in in-line probing buffer for 40 h at RT. The reaction products were then separated by gel electrophoresis and visualized

by phosphoimaging. The results are depicted in **Fig. 2.7 A**. In-line probing of the 64 nt long minimal region confirmed the existence of the two hairpins predicted by Dynalign. *In vitro*, both stem-loop structures form a 6 bp long stem with an A-U closing base pair that is capped by a terminal UAU tri-nucleotide loop (**Fig. 2.7B**). Notably the stem of hairpin 2 is only composed of A-U base pairs. The basal U-A base pair of hairpin 3 is not formed *in vivo*. Therefore, both hairpins have an identical stem length of 6 bp. Furthermore, a small stem of 5 bp is formed by the surrounding sequences. Sequence comparison of the 64 nt long *UCP3* minimal motif with the corresponding sequences from 46 other mammals annotated in the UCSC Genome Browser showed that both RNA hairpins are highly conserved in primary sequence compared to the surrounding sequences (**Supp. Fig. 5.5**), further underlining their importance in gene regulation. To confirm the hypothesis that the two RNA hairpins are crucial for regulation, a screening was performed to identify the exact areas necessary for post-transcriptional gene regulation by the 64 nt long element. Therefore, the sequence was mutated in blocks of seven nucleotides to balance the number of clones and sequence perturbation. To a achieve maximum variation, A to C and G to U and *vice versa* were mutated in the individual blocks. This not only changed the sequence but also all pyrimidine to purine bases and *vice versa*. All mutations that prevented structure formation of the hairpins led to a significant increase in luciferase activity. Even the mutant 8-14, which shortened the stem of hairpin 2 by only 1 bp down to 5 bp, already showed a significantly higher luciferase activity compared to the wt sequence (**Fig. 2.7C**). In contrast, mutants which only changed the surrounding sequences had no significant effect on luciferase activity. These results also fit perfectly to the conservation of the primary sequence. In summary, I was able to show that the 64 nt long minimal motif folds into a tandem RNA hairpin that largely corresponds to the structure predicted by *Dynalign* and that both hairpins are essential for the regulatory function of the element *in vivo*.

2.1.5 The *UCP3* Element Reduces mRNA Half-Life

Cis-regulatory elements can reduce gene expression by either reducing translation efficiency or mRNA half-life. Regulation of mRNA turnover in the cytoplasm is important for controlling the abundance of cellular transcripts and, in turn, protein abundance. To investigate whether the *UCP3* element reduces gene expression by reducing the RNA half-life, I investigated the RNA half-life of RNAs encoding either the 64 nt wt motif or a double mutant that prevented the formation of both hairpins (MUTI/II). Therefore, a dual fluorescence reporter plasmid was cloned, containing an mCherry and a GFP reporter under the control of a shared CMV promoter/enhancer element. The plasmid also contained an FRT site for genomic integration into a FLPin host cell line. *UCP3* wt and MUTI/II were fused to the open reading frame of GFP and stably integrated into the genome of HeLa cells. To show that the genomic integration was successful, both GFP and mCherry levels were quantified by flow cytometry (**Fig. 2.8A**). Normalized GFP levels showed a 7.7-fold reduction in GFP levels by the *UCP3* wt element compared

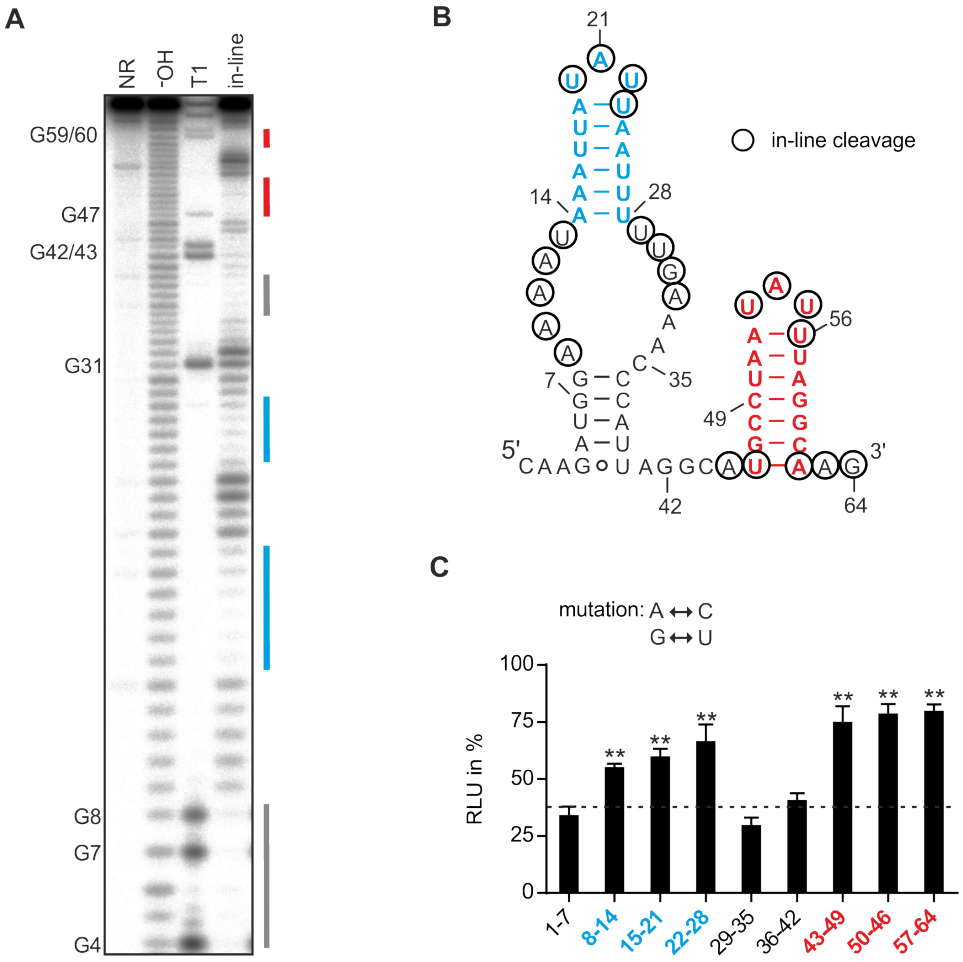

Figure 2.7: The *UCP3* wt element folds into two hairpins. **(A)** In-line probing analysis of *UCP3* wt RNA. The RNA was loaded directly (NR, no reaction), subjected to cleavage by RNase T1 or alkaline hydrolysis, or incubated for 40 h at room temperature and pH 8.3 (in-line) prior to Urea PAGE. Paired regions are indicated by identically colored lines. **(B)** Predicted lowest free energy secondary structure of the *UCP3* wt element by RNAstructure. Nucleotides detected by in-line probing shown in (A) are circled. **(C)** Luciferase activity of *UCP3* mutants for the identification of motifs essential for gene regulation. Adenine was mutated to cytosine, guanine to uracil and *vice versa*. Numbers indicate mutated nucleotide positions. Firefly luciferase activity was normalized to *Renilla* luciferase as internal transfection control. Values are normalized to an empty vector control, without *UCP3* 3'UTR sequences. Luciferase activity of the *UCP3* wt element is indicated as dashed line. n = 3. (**) P-value < 0.01. Adopted and modified from [113].

to MUTI/II (**2.8B**). These values confirmed the results already obtained in luciferase assays and underline the self-sufficiency of the repressive element. To determine whether the reduction was a consequence of a decreased *GFP* mRNA abundance, the amount of mRNA was quantified by RT-qPCR. A corresponding reduction could be observed at the mRNA level (**Fig. 2.9A**). However, changes in mRNA abundance are not necessarily

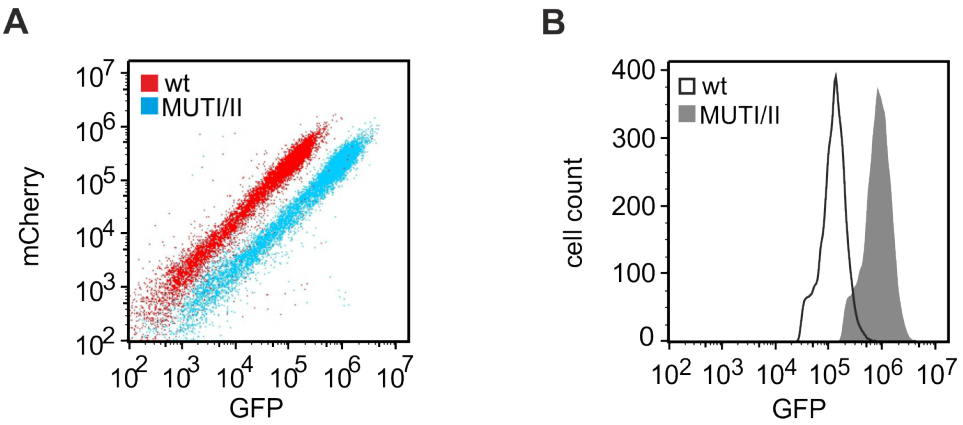

Figure 2.8: Genomic integration of *GFP-UCP3* wt element fusion mRNAs. (A) GFP and mCherry fluorescence of *UCP3* wt and double mutant (MUTI/II). *GFP-UCP3*-fusion constructs were stably integrated into the genome of HeLa cells. GFP and mCherry fluorescence were measured by flow cytometry. n = 3. **(B)** GFP fluorescence of *UCP3* wt and double mutant (MUTI/II). GFP fluorescence was measured by flow cytometry and was normalized to mCherry values. n = 3. Adopted and modified from [113].

caused by alterations in mRNA stability. For example, mRNA biogenesis in the nucleus such as transcription, RNA processing, and/or mRNA export may be altered by the introduction of the *UCP3* wt element into the 3'UTR of an mRNA. Thus, alterations in the steady-state level of mRNA may not reflect changes in mRNA stability [143]. A relatively simple way of analyzing mRNA kinetics involves blocking cellular transcription with inhibitors like actinomycin D. This antibiotic interferes with transcription by intercalating into the DNA. The amount of a particular mRNA remaining at various time points after the treatment is then used to calculate the mRNA decay rate. Determination of mRNA half-lives after treatment with actinomycin D showed that *GFP* mRNA containing the *UCP3* wt element in its 3'UTR is very unstable (half-life of $\sim$ 1.5 h) compared to the double mutant, for which no reduction of mRNA levels was detectable over the time course of the experiment (**Fig. 2.9B**). In summery, I could show, that the repression of gene expression by the *UCP3* 3'UTR element is the result of a significant reduction in mRNA half-life compared to the control mRNA. Summarizing this part of my thesis, I successfully identified a structually conserved regulatory element in the 3'UTR of the *UCP3* mRNA. This element was predicted by Dynalign to be structurally conserved across among human and mouse. I identified a 64 nt minimal motif that is sufficient and necessary for regulation. Structural probing revealed that the sequence

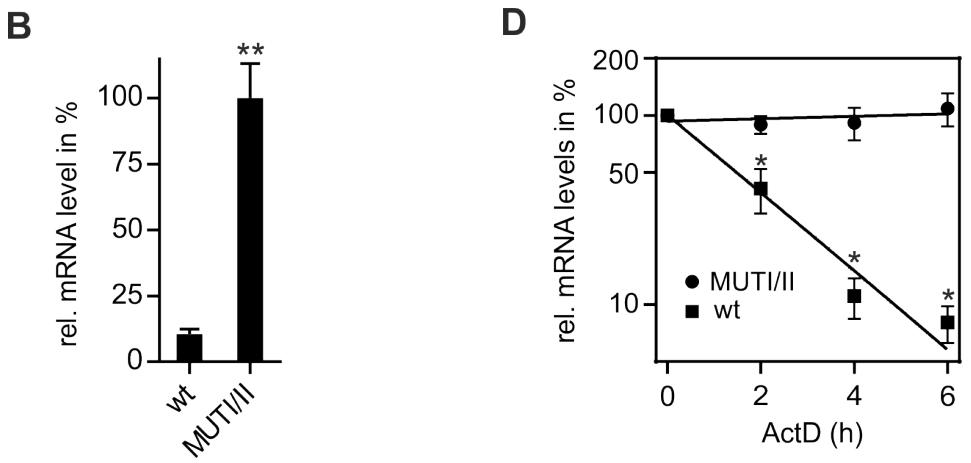

Figure 2.9: mRNA half-life of *UCP3* wt and MUTI/II-containing mRNAs. (A) RT-qPCR quantification of *GFP* mRNAs containing the *UCP3* wt element (wt) or double mutant (MUTI/II). Total RNA from HeLa cells stably expressing one of the two constructs was isolated and *GFP* mRNA levels quantified by RT-qPCR. **(D)** HeLa cells stably expressing one of the two constructs were treated with 5 µg/µl actinomycin D (ActD). Thereafter, total RNA was isolated at 2 h intervals and *GFP* mRNA levels quantified by RT-qPCR. *GFP* values are normalized to the housekeeping gene *RPLP0*. n = 3. (**) P-value < 0.01 (*) P-value < 0.05. Adopted and modified from [113].

adopts the tandem stem-loop structure predicted by Dynalign and that both hairpins are necessary for robust regulation. Furthermore, I was able to show that repression of gene expression by the *UCP3* wt element is a direct consequence of a significant reduction in mRNA stability.

2.2 Identification of *Trans*-acting Factors

RNA affinity purification is a powerful tool, which allows for the identification and characterization of various proteins bound to a certain RNA. There are a variety of different strategies that are based on either covalent or non-covalent immobilization of the target RNA to an affinity matrix. The high affinity binding of biotin by avidin has made the biotin-avidin association an extremely powerful tool for affinity chromatography and a method of choice for many researchers. With this method ITAFs which modulate XIAP and cIAP1 IRES activity were successfully identified and characterized [144]. In this part of my thesis, I used RNA affinity to specifically enrich *UCP3* wt RNA-binding proteins. In depth mass spectrometry analysis identified the proteins Roquin-1 and Roquin-2 as *trans*-acting factors that bind to the *UCP3* wt element. Detailed *in vivo* and *in vitro* analysis confirmed that Roquin binds to the *UCP3* wt element, and thereby reduces expression of *UCP3* wt element encoding genes. RNA mobility shift assays revealed that both RNA hairpins in the *UCP3* wt element are bound individually by Roquin

through its core ROQ domain. These stem-loops are similar to known Roquin binding sites, so called CDEs, in that they code for a 6 bp long stem capped by a tri-nucleotide loop. Both stem-loops are required for efficient regulation by the *UCP3* wt element. In addition, I could show that endogenous *UCP3* mRNA is repressed by Roquin in murine C2C12 cells.

2.2.1 RNA Affinity Purification of RNA-binding Proteins

One major goal of this project was the identification of *UCP3* wt binding proteins since functional interaction partners have not been previously described. To identify RBPs that specifically bind to my RNA of interest, I performed an optimized mRNP purification based on the high affinity biotin-avidin interaction. Thus, I transcribed *UCP3* wt RNA as well as mutants preventing formation of the first or second hairpin (MUTI; MUTII) or both (MUTI/II) *in vitro*. Additionally, a randomized sequence with equal nucleotide composition was synthesized (**Fig. 2.10A**). Prior to RNA affinity purification all RNAs constructs were tested for their regulatory properties in HEK293 cells (**Fig. 2.10B**). As expected individual mutation of either RNA hairpin led to a significant in-

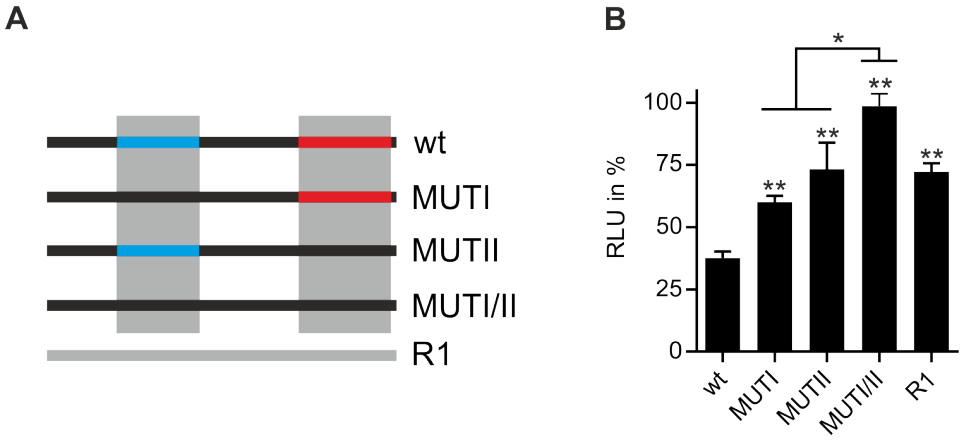

Figure 2.10: *UCP3* RNAs used for RNA affinity purification. (A) Overview of *UCP3* constructs used for RNA affinity purification.**(B)** Luciferase activity of *UCP3* constructs and a randomized control with the same nucleotide composition (*R1*) used for RNA affinity purification. Firefly luciferase activity was normalized to *Renilla* luciferase as internal transfection control. Values are normalized to an empty vector control. n = 3. (**) P-value < 0.01. (*) P-value < 0.05. Adopted and modified from [113].

crease of luciferase activity compared to the *UCP3* wt element. Interestingly mutation of both hairpins led to a further significant increase in luciferase activity compared to the single mutants, whereas randomization of the element (*R1*) lead to a similar increase in luciferase activity as the single mutants. Next, the *in vitro* transcribed RNAs were 5'-biotinylated and coupled to steptavidin-conjugated magnetic beads for affinity purification of proteins from HEK293 cells. After stringent washing, RNA-bound proteins

were eluted and separated by SDS-PAGE (**Suppl. Fig. 5.8**). While the majority of proteins bound to all five RNAs, I was able to identify differentially enriched proteins by mass spectrometry analysis based on unique peptide counts. 229 proteins were detected by mass spectrometry using a cut-off of at least five unique peptides detected in at least one of the samples. While most purified proteins associated with all five RNAs, ten of these proteins were not detected in the two controls, i.e. the double mutant and the randomized sequence (**Tab. 2.1**). Among these ten, the paralogs Roquin-1 and Roquin-2

Table 2.1: Proteins enriched by RNA affinity purification. Adopted and modified from [113].

			No. of peptides				
Description	**Gene Symbol**	**MW [kDa]**	**wt**	**MUT1**	**MUTII**	**MUTI/II**	**R1**
Roquin-2	RC3H2	132	30	20	23	0	0
Zinc finger CCCH domain-containing protein 7A	ZC3H7A	111	19	0	29	0	0
Spermatid perinuclear RNA-binding protein	STRBP	72	14	10	11	0	0
Roquin-1	RC3H1	126	7	1	6	0	0
Gem-associated protein 5	GEMIN5	168	5	26	0	0	0
ATP-dependent RNA helicase DDX1	DDX1	74	5	12	15	0	0
RNA-binding protein MEX3A	MEX3A	54	3	0	5	0	0
tRNA-splicing ligase RtcB homolog	RTCB	55	4	4	8	0	0
RNA-binding protein 47	RBM47	64	2	0	8	0	0
RNA-binding protein Raly	RALY	32	3	5	0	0	0

were highly enriched. Roquin-1 has previously been described to destabilize its target mRNA coding ICOS to prevent autoimmunity [145] [51]. It has since been found to play an essential role in controlling the levels of mRNAs coding for various key proteins involved in immune regulation [57]. In addition, Roquin destabilizes *TNF* mRNA by binding to a small structured RNA motif, a so-called CDE, in the *TNF* 3'UTR [56]. Thus, Roquin proteins appeared to be promising candidates to regulate gene expression by binding to the *UCP3* wt element. The binding of Roquin proteins to the *UCP3* wt element was verified by western blot analysis after RNA affinity purification. Confirming mass spectrometry results, I was able to detect Roquin association with *UCP3* wt RNA as well as the single mutants, where formation of one of the hairpins was prevented. In contrast, no binding was detected with the double mutant (**Fig. 2.11**).

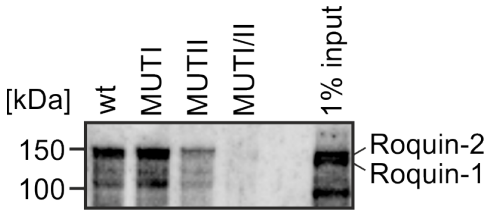

Figure 2.11: RNA affinity purification of Roquin-1/-2 with different *UCP3* RNAs. For RNA affinity purification HEK293 whole cell lysates were incubated with the different *UCP3* RNAs **2.10(A)**. Roquin-1 and Roquin-2 were visualized by Western Blot using anti-Roquin antibody. n = 2. Adopted and modified from [113].

2.2.2 Roquin Proteins Mediate Repression by Binding the *UCP3* wt Element

In the previously described experiments, I could show that the *UCP3* wt element is bound by Roquin *in vitro*. To determine whether Roquin leads to the repression of gene expression by binding to the *UCP3* wt element, I performed transient knockdown of Roquin proteins and quantified the influence on luciferase activity. HEK293 cells were transiently transfected with siRNAs targeting either Roquin-1, Roquin-2 or both. 24 h post siRNA transfection the *UCP3* wt element-encoding luciferase reporter plasmid was transfected to investigate the influence of the Roquin knockdown on luciferase activity. Successful knockdown of Roquin was verified by western blot analysis (**Fig. 2.12A**). Knockdown of Roquin-1 led to a small, but significant increase in luciferase activity

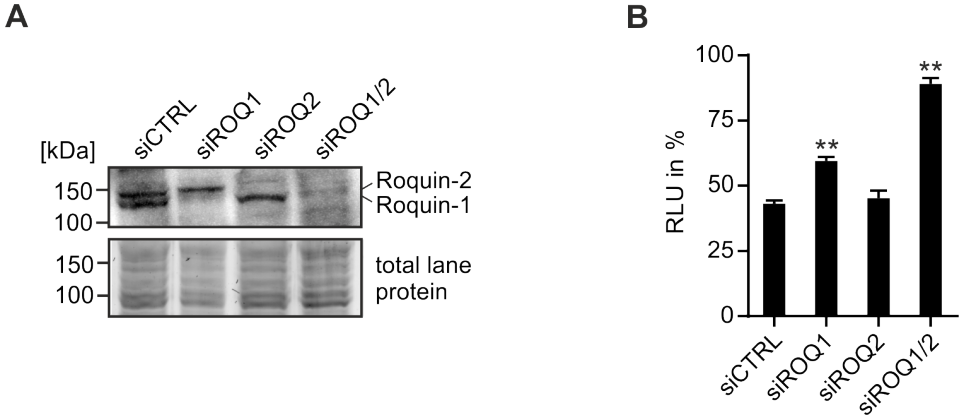

Figure 2.12: Roquin regulates gene expression via the *UCP3* tandem CDE. (A) Western Blot of Roquin-1 and Roquin-2 after siRNA-mediated knockdown. Anti-Roquin was used to verify the respective knockdown. Total lane protein is shown as loading control. n = 3. **(B)** Luciferase activity of the *UCP3* wt element after siRNA-mediated knockdown of Roquin proteins. Firefly luciferase activity was normalized to *Renilla* luciferase as internal transfection control. Values are normalized to an empty vector control, without *UCP3* 3'UTR sequences. n = 3. (**) P-value <0.01. (*) P-value <0.05. Adopted and modified from [113].

compared to the control. No influence on luciferase activity could be observed with the knockdown of Roquin-2. Full restoration of luciferase activity was achieved only by simultaneous knockdown of Roquin-1/-2 (**Fig.2.12B**). This finding is in line with the previously reported redundant function of Roquin-1/-2 in repression of gene expression [53]. In addition, transient knockdown of Roquin-1/-2 in HeLa cells stably expressing *UCP3* wt encoding *GFP* mRNA (see section **2.1.5**) led to a significant increase in GFP levels **Suppl. Fig.5.9A and B**.

So far, I only demonstrated repression by the *UCP3* wt element in a reporter gene context. To investigate whether endogenous *UCP3* is a Roquin target, *UCP3* mRNA

was quantified via RT-qPCR. UCP3 is predominantly expressed in smooth skeletal muscle and brown adipose tissue [146] [147] [148] and only detectable in trace amounts in HEK293 cells. To assess changes in *UCP3* mRNA abundance after Roquin knockdown, I performed a transient knockdown of mouse Roquin-1/-2 in C2C12 murine myoblasts. After siRNA transfection, cells were grown for 48 h to reach confluency and differentiation induced by adding growth medium containing 2% horse serum. 24 h post induction, protein and RNA samples were prepared. Efficient knockdown 72 h post siRNA transfection was verified by western blot analysis (**Fig. 2.13A**) and quantification of Roquin-1 and Roquin-2 mRNA levels (**Fig. 2.13B**). The *UCP3* mRNA level was significantly higher (approx. 3-fold) in Roquin-1/-2 knockdown cells compared to the control (**Fig. 2.13C**). Overall these results confirmed the role of Roquin in post-transcriptional regulation by the *UCP3* wt element.

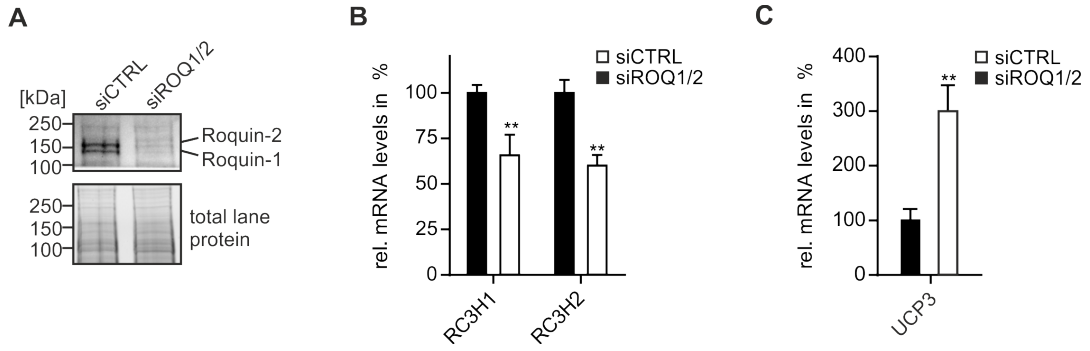

Figure 2.13: Roquin knockdown reduces *UCP3* mRNA levels in C2C12 cells. (A) Western blot of Roquin-1 and Roquin-2 after siRNA-mediated knockdown in C2C12 cells. AntiRoquin was used to verify the respective knockdown. Total lane protein is shown as loading control. (**B** and **C**) RT-qPCR quantification of (**B**) *RC3H1* and *RC3H2* and (**C**) *UCP3* mRNA levels after siRNA-mediated knockdown of Roquin-1 and Roquin-2 in C2C12 cells. Values are normalized to the housekeeping gene *GAPDH*. n = 3. (**) P-value < 0.01. Adopted and modified from [113].

2.3 Redefining Roquin Binding Preferences

In the previous chapter, I explained in detail how Roquin represses gene expression by binding to the *UCP3* wt element. In this chapter, I describe how I analyzed the biochemical properties of the *UCP3* wt-Roquin interaction. In RNA mobility shift assays, I could show that Roquin binds to the *UCP3* wt RNA with higher affinity compared to the mutants, where formation of one of the two CDEs was prevented. Mutation of the core ROQ domain results in a loss of binding activity. Roquin affinity to the *UCP3* wt RNA is comparable to known targets like *TNF*, *ICOS* and *Ox40*. Since both CDEs in the *UCP3* wt element deviate from the published CDE consensus, derived

from the mutational analysis of the *TNF* CDE, I performed an extensive mutational analysis of the *UCP3* wt element. The results of these experiments substantiate *in vitro* data, that the Roquin recognition is almost exclusively shape dependent and allowed me to formulate a new CDE consensus that is more relaxed than the previously proposed CDE consensus. This also suggests that many more genes could encode functional CDEs and thus are regulated directly by Roquin.

2.3.1 The *UCP3* Element Encodes Two Roquin Binding Sites

In the RNA affinity purification experiments, Roquin could be purified with *UCP3* MUTI and MUTII RNA, but not the double mutant (see **Tab. 2.1** and **Fig. 2.11**), thus suggesting individual binding of Roquin to both CDEs in the *UCP3* wt element. However, mutation of one of the two CDEs was sufficient to significantly increase luciferase activity (see **Fig. 2.10B**). To investigate this discrepancy in more detail, I performed RNA gel retardation assays with purified Roquin-1 protein and *UCP3* wt RNA as well as single or double mutants of both CDEs. To minimize the effect on overall folding of the RNA molecules, only single nucleotide mutations were introduced. The mutations only effected the closing base pair and prevented triloop formation (**Fig. 2.14A**). Overall folding was assessed by RNA secondary prediction by RNAstructure

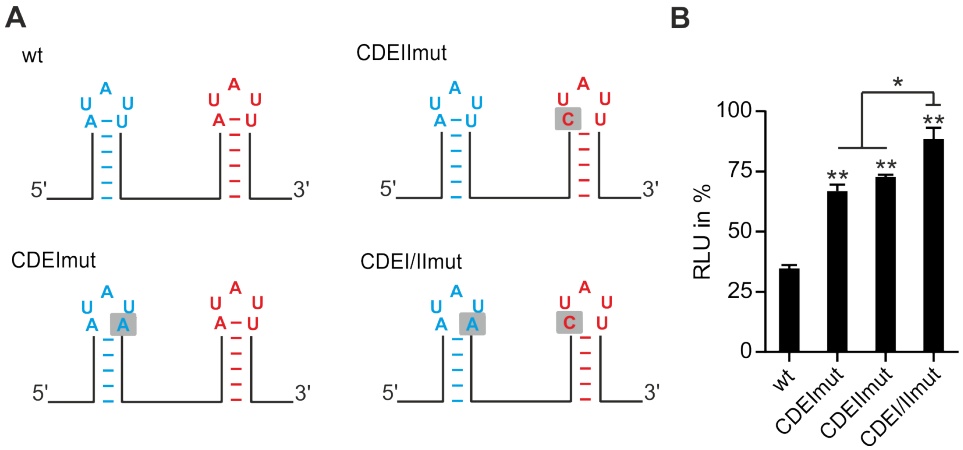

Figure 2.14: Overview of *UCP3* constructs used for binding experiments. (A) Schematic overview of RNAs used in EMSA experiments. **(B)** Luciferase activity of *UCP3* constructs shown in **(A)**. Firefly luciferase activity was normalized to *Renilla* luciferase as internal transfection control. Values are normalized to an empty vector control, without *UCP3* 3'UTR sequences. n = 3. (**) P-value < 0.01. (*) P-value < 0.05. Adopted and modified from [113].

(**Suppl. Fig. 5.10**). In accordance with my previous results, individual mutation of CDEI or CDEII to a pentaloop increased luciferase activity to $\sim 67-72\%$ compared to the empty vector control. Mutation of both CDEs led to a further small, but significant

increase in luciferase activity to ∼ 89% of the control (**Fig. 2.14B**). These findings suggested, that Roquin can bind to both CDEs individually, but that both CDEs are necessary for robust regulation by the *UCP3* wt element.

Roquin-1 is a multi-domain protein. The Roquin domain organization is depicted in **Fig. 2.15** (Roquin-1 fl). Roquin proteins share a highly similar N-terminus which consists of a RING domain, that probably functions as an E3 ubiquitin ligase and two RNA-binding domains: the ROQ domain, which is characteristic of Roquin proteins and a CCCH zinc finger. The C-terminus of the protein is largely unstructured and contains a proline-rich domain and a coiled-coil domain. Previous studies have shown that Roquin binds the *TNF* CDE through its ROQ domain. Mutation of the ZnF as well as deletion of the C-terminus had no effect on the association of Roquin with the *TNF* CDE [56]. Biochemical studies investigating the binding of Roquin to composite RNA elements in

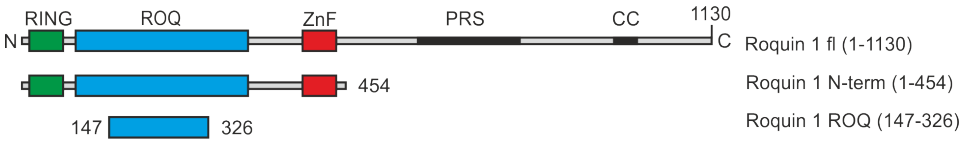

Figure 2.15: Roquin-1 domain organization. Domain organization of mouse Roquin-1 and overview of Roquin fragments used for binding experiments. fl: full length protein; N-term: Roquin-1 N-terminus including the RING, ROQ and ZnF domain; ROQ: Roquin-1 core ROQ domain. Figure adopted from [59]. Adopted and modified from [113].

the 3'UTR of *ICOS* [58] and *Ox40* [64] mRNA have previously been described. In these examples ADEs and CDEs function in conjunction to achieve efficient Roquin binding. In order to dissect Roquin binding to the *UCP3* wt element, I performed EMSAs with *in vitro* transcribed RNAs, equivalent to *UCP3* mutants depicted in **Fig. 2.14**, and purified Roquin-1 ROQ domain (**Fig. 2.16A**) and the complete Roquin N-terminus (**Fig. 2.16C**). Therefore the RNA was radioactively labeled at its 5' end and incubated with increasing amounts of protein. The formed complexes were than separated by native PAA electrophoresis in order to detect differences in RNA migration. The EMSA experiments showed that both, the core ROQ domain as well as the complete N-terminus bind with high affinity to the *UCP3* wt element with respective apparent K_D values of $138 \pm 2\,\text{nM}$ (ROQ) and $225 \pm 1\,\text{nM}$ (N-term). Both Roquin-1 fragments showed a significantly lower affinity to the two single mutants (**Fig. 2.16B** and **D**). For the double mutant no binding could be observed. Interestingly in the binding reactions with the Roquin N-terminus and *UCP3* wt two complexes were detected in the gel which is consistent with the fact that two Roquin entities can bind to the tandem CDE. In contrast, only one band was detected upon binding of the ROQ domain with *UCP3* wt. However, the migration of this band was less retarded with the single mutation of the CDEs compared to the *UCP3* wt. This suggests individual binding of two core ROQ domains to the *UCP3* wt RNA.

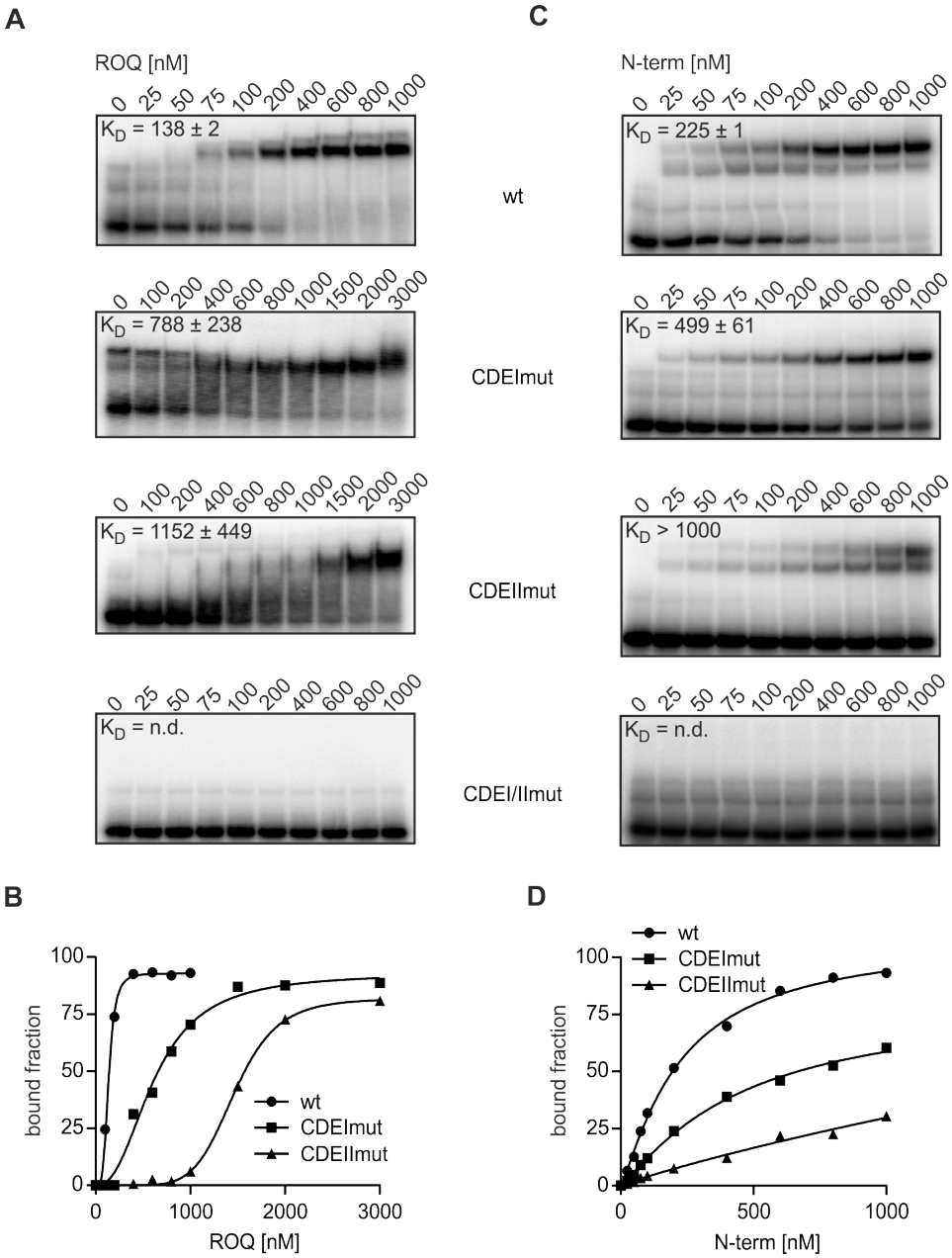

Figure 2.16: Both *UCP3* CDEs are required for efficient Roquin binding. Binding of recombinant Roquin-1 to the *UCP3* constructs shown in **Fig. 2.14**. Radiolabeled RNAs were incubated with increasing amounts of Roquin-1 ROQ domain **(A)** or N-terminus **(C)**. The apparent dissociation constant (K_D) was calculated from 2-3 independent experiments. (**B** and **D**) Representative quantification of EMSA experiments with Roquin-1 ROQ domain (**B**) or N-terminus (**D**). n.d. = not determined. Adopted and modified from [113].

For further analysis of the reaction stoichiometry, I performed a stoichiometric binding experiment. In this EMSA experiment the protein is titrated to RNA of known concentration with trace amount of radiolabeled RNA for detection. Total RNA concentration must be > 50-fold greater than the K_D. By plotting the bound fraction of RNA by the molar equivalent of protein a saturation curve is obtained. This curve can than be compared to theoretical saturation curves and the binding stoichiometry determined [149]. Stoichiometric binding experiments confirmed that *UCP3* wt RNA is bound by two core ROQ domains (**Fig. 2.17A**). This finding is in accordance with the simultaneous detection of both CDEs by the core ROQ domain. Previous studies showed that mutation of the core ROQ domain inhibits Roquin binding to the *TNF* CDE. As demonstrated with the *TNF* CDE, no binding was observed in EMSAs with triple mutant ROQ domain (K220A, K239A, R260A) and UCP3 wt RNA verifying that Roquin binding to the *UCP3* wt element is mediated by its core ROQ domain.

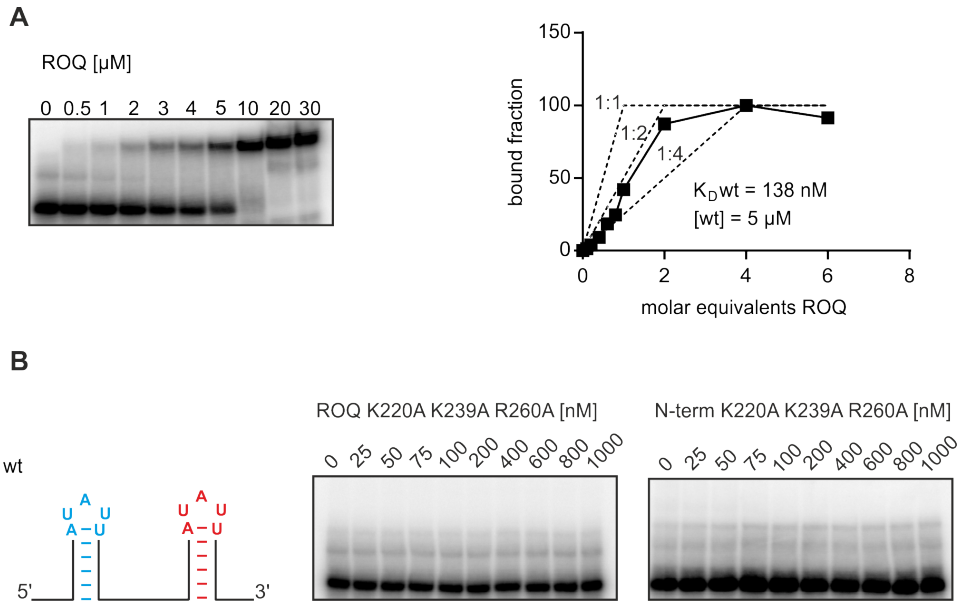

Figure 2.17: The core ROQ domain is essential for binding of Roquin-1 to the *UCP3* wt element. (A) Left: Increasing amounts of Roquin-1 ROQ domain were incubated with 5 µM *UCP3* wt RNA containing a trace of radioactively labeled RNA. Right: Comparison of the observed binding saturation with theoretical saturation curves for a 1:1; 1:2 and 1:4 binding stoichiometry. n = 2. **(B)** Radiolabeled *UCP3* wt RNA was incubated with increasing amounts of triple mutant (K220A, K239A, R260A) Roquin-1 ROQ domain or N-terminus. n = 2. Adopted and modified from [113].

2.3.2 *In vivo* Mutational Analysis of the *UCP3* Tandem CDE

Since the discovery of the CDE in the 3'UTR of *TNF*, a series of *in vitro* and *in vivo* mutation studies have been performed with the aim to identify the structure and sequence

requirements Roquin has for an effective binding to a CDE. Leppek *et al.* investigated the influence of mutations that either altered the RNA hairpin structure or its sequence on the mRNA half-life of a ß-globin reporter gene [56]. Based on their data, they concluded a restrictive consensus structure that had specific sequence requirements for the apical part of the stem (see **Fig. 3.1A**). Based on the findings of Leppek *et al.*, a number of structural studies were performed to investigate the *in vitro* binding specificity of Roquin to the *TNF* CDE. Schlundt *et al.* and Codutti *et al.* could show that the *in vitro* binding of Roquin to the *TNF* CDE is almost exclusively shape dependent. Their data showed only a very low sequence dependency and allowed to formulate a relaxed consensus. This consensus also included the meanwhile published findings of Janowski *et al.* which showed that Roquin is also able to bind U-rich RNA hexaloops, so-called alternative decay elements (ADE). The proposed consensus consists of an RNA hairpin with a stem length of 6 bp, whereby the upper two base pairs must each be pyrimidine-purine (Y-R) (**Fig. 3.1B**) [57]. However, the proposed CDE consensus was never subjected to a detailed *in vivo* analysis.

Interestingly, both CDEs found in the 3'UTR of *UCP3* deviate from the *TNF* CDE derived consensus as well as from the more general proposed consensus which includes the *in vitro* structural analysis. In order to clearify the binding preferences for Roquin *in vivo*, I performed an in depth mutation screen to evaluate the contribution of the different features that have previously been described as being important to Roquin based regulation. To assess the importance of different features in the *UCP3* wt element I introduced mutations into CDEI, CDEII or both and measured the effect on luciferase activity. An overview of all mutations is depicted in (**Fig. 2.18A**). First of all I investigated the influence of the apical base pair of the stem which closes the tri-nucleotide loop. In the published CDE consensus, the polarity of the closing pair is decisive for the binding of Roquin. Mirroring the upper two Y-R base pairs of the *TNF* CDE results in a significant increase of the β-globin reporter mRNA half-life [56]. In contrast to the *TNF* CDE, both *UCP3* CDEs feature an R-Y closing base pair. Mutation of the closing base pair to any other Watson-Crick base pair in either CDEI, CDEII or both did not influence luciferase activity (**Fig. 2.18B**). Mutation of the closing base pair into a G∘U or U∘G wobble base pair however, led to a significant increase in luciferase activity (**Suppl. 5.11**). However it did not increase the luciferase activity to the same extend as the pentaloop mutations (**Fig. 2.14B**). This suggests that apical wobble base pairs might occur in endogenous Roquin binding sites.

In the next step, I wanted to investigate whether the stability of the stem has an influence on Roquin binding *in vivo*. Therefore, I stabilized CDEI, which consists exclusively of A-U base pairs, by introducing one (M10) or two G-C (M11) base pairs. Increasing the ΔG of the CDEI stem did not influence luciferase activity (**Fig. 2.18C**). Smiliary, destabilizing CDEII by exchanging the central G-C base pair to an A-U base pair (M13) also had no effect on luciferase activity. In addition, I completely exchanged CDEI with CDEII which had also no effect on luciferase activity *in vivo* compared to the *UCP3* wt element (**Suppl. 5.11**).

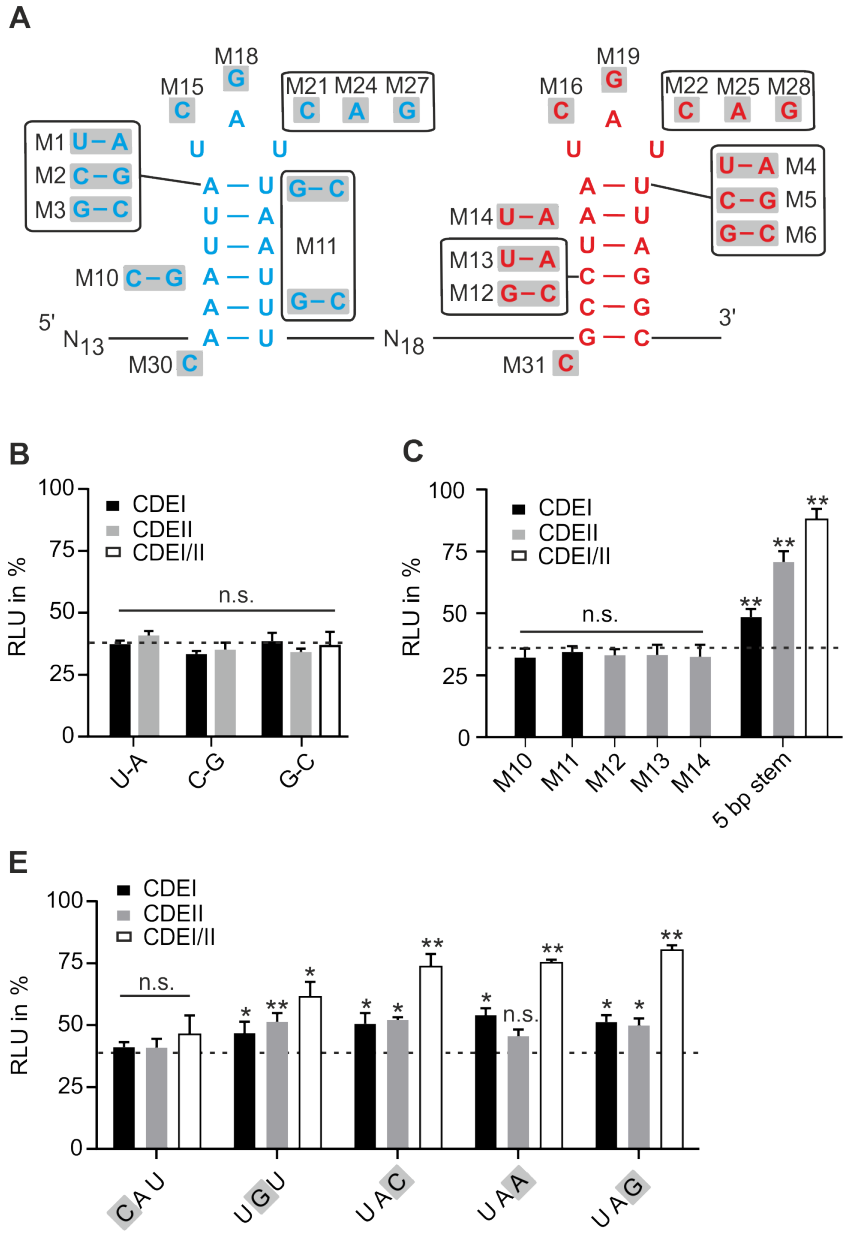

Figure 2.18: Mutational analysis of the *UCP3* tandem CDE. (A) Overview of UCP3 mutants. **(B)** Luciferase activity of closing base pair mutants. n = 3. **(C)** Luciferase activity of stem mutants. n = 3. **(D)** Luciferase activity of triloop mutants. n = 3. (**B-E**) Firefly luciferase activity was normalized to *Renilla* luciferase as internal transfection control. Values are normalized to an empty vector control, without *UCP3* 3'UTR sequences. Luciferase activity of the *UCP3* wt element is indicated as dashed line. (**) P-value < 0.01. (*) P-value < 0.05. n.s. = not significant. Adopted and modified from [113].

In the previously performed hepamerscreen (see **Fig. 2.7C**) mutation of nucleotides 8-14, which shortens the stem of CDEI by only one base pair down to 5 bp, already significantly restored luciferase activity. To exclude the possibility that mutation of the surrounding nucleotides contributed to this effect, I introduced point mutations (M30, M31, M32) to shorten the stem of CDEI, CDEII or both. Shortening of the CDE stem to 5 bp significantly increased luciferase activity, verifying that Roquin binding is dependent on CDE stem length *in vivo*. Conversely, it has been shown in the *TNF* CDE mutagenesis screen that extending the stem to 9 bp also inactivated CDE-mediated mRNA decay [56].

The results of the *TNF* CDE mutagenesis also suggested that a purine stack of at least three consecutive purines at the 3' side of the stem is important for CDE mediated mRNA decay. This assumption has already been questioned by subsequent studies *in vitro* [57], but has never been thoroughly investigated *in vivo*. I could demonstrate that neither the introduction of a purine stack into CDEI (M11) nor destruction of the putative purine stack in CDEII (M12) had a significant influence on luciferase activity. Taken together, this suggested that Roquin binding to both CDEs in the *UCP3* wt element is not dependent on stem stability or sequence nor is the polarity of the closing base pair or the presence of a purine stack on the 3' side of the stem a prerequisite for Roquin binding *in vivo*.

Previous *in vitro* studies suggested Y-R-N as a possible motif for the tri-nucleotide loop [60] [57]. However, this motif has not yet been comprehensively tested *in vivo*. To investigate the influence of the loop nucleotides on Roquin binding *in vivo*, I mutated nucleotides 1-3 of the loop to the corresponding other possible nucleotides and quantified luciferase activity (**Fig. 2.18D**). The Mutation of U at position 1 to a C in CDEI, CDEII or both had no significant influence on the luciferase activity. This finding is in line with results from previous studies with the *TNF* CDE. On the other hand, the mutation of A at position 2 to G led to a significant increase in luciferase activity compared to the *UCP3* wt element. The effect is most pronounced in the double mutant. This result is in contradiction to the *TNF* CDE, which endogenously encodes for a G at this position. Also mutation of G to A does not affect gene repression by the *TNF* CDE [56]. The mutation of U in position 3 of the triloop to any other nucleotide significantly increased luciferase activity. This was especially pronounced in the double mutants. While luciferase activity in double mutants increases to $\sim 70 - 75\%$, it does not reach the level of pentaloop mutants (90%) suggesting that though Roquin binding is impaired, there is binding activity remaining. From the data shown, I concluded that -YAU- triloops are most effective for the regulation of gene expression by Roquin. This motif also corresponds to the conservation of the *UCP3* wt sequence in mammals where -YAU- occurs almost exclusively (**Suppl. 5.5**). However, it cannot be excluded that endogenous Roquin targets code for a -YRN- motif, since these sequences still have a certain regulatory activity. In summary, the presented *in vivo* data set let me formulate a new CDE consensus (**Fig. 2.19C**). The *UCP3* wt derived relaxed consensus consists of a 6-8 bp long stem capped by a Y-R-N tri-nucleotide loop. The length of the stem

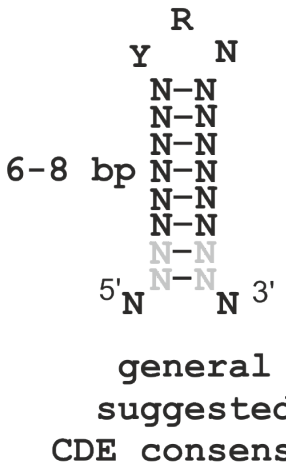

Figure 2.19: Roquin recognizes a relaxed CDE consensus. A new general suggested, relaxed CDE consensus based on the *in vivo UCP3* wt mutational analysis presented in this work.

is based on my results that shortening the stem to 5 bp already induced a significant increase in luciferase activity and that extension of the *TNF* CDE to a 9 bp stem leads to complete inactivation of CDE-mediated mRNA decay. Taken together, this suggests that Roquin binding is mainly shape-dependent with only very few requirements for the loop sequence. This supports previous structural studies that showed a strong shape dependency of Roquin binding. Besides triloop nucleotide composition, stem length also seems to be a very important feature of Roquin mediated gene repression. Further, the above presented results suggest that many more genes may encode for functional Roquin binding sites.

2.4 Identification of New High Affinity Targets of Roquin

As described in the previous chapter, the detailed *in vivo* mutational analysis allowed me to suggest a new relaxed CDE consensus. In this part of my thesis this data was used to perform a computational analysis of the human genome for the identification of new Roquin targets. Leppek *et al.* performed a descriptor search based on their initially proposed *TNF*-derived CDE consensus. They identified 109 CDEs in 108 genes. 56 of the predicted CDEs were highly conserved among mammals. To verify that CDE-containing mRNAs are targets of Roquin, they performed an immunoprecipitation (IP) of Roquin-1 followed by RNA sequencing. The most enriched RNA they identified was *NFKBID* which contains a tandem CDE in its 3'UTR. From that they concluded that CDE-containing genes are primary targets of Roquin. However, it must be said that

only 15 of all 95 mRNAs significantly enriched in the Roquin IP contained a conserved CDE, corresponding to the *TNF*-derived descriptor, in their 3'UTR. This small overlap suggests that their descriptor based computational prediction was not able to reliably predict most binding site containing genes. With the help of the data generated in this thesis and the general CDE motif derived from it, it should be possible to predict many more genes that code for a functional CDE and therefore be primary targets of Roquin.

2.4.1 Computational Prediction of CDEs

In collaboration with the Mathews lab (Rochester University, NY), we predicted CDE-containing genes on a genome wide scale. Therefore all human 3'UTRs were searched for potential CDEs based on the *UCP3* derived CDE consensus(**Fig. 2.19**). For this search G∘U wobble base pairs were excluded, because this was the only mutation that significantly increased luciferase activity. For the descriptor search, all human 3'UTR sequences were obtained from the UCSC genome browser (hg19). The program RNAmotif was run to identify CDE-containing genes. The program takes an RNA descriptor, an RNA sequence file and outputs the positions where the RNA motif is found. Using RNAmotif $\sim$ 2600 genes that contained at least one predicted CDE were identified. As we have seen in the previous results that the correct folding of the CDE is decisive for its correct function in gene repression, the estimate folding probability of the predicted CDEs was calculated. In order to calculate the estimate folding probability a C++ program called ProbStemloop was written, using RNAstructure. ProbStemloop takes three input parameters: the sequence file, the starting position of a stem-loop and the length of the stem-loop and outputs the folding probability of the the motif. The sequences processed by ProbStemloop were truncated to 400 nt upstream and 400 nt downstream, as has been demonstrated that this much sequence is a sufficient representation of the local folding in mRNA [150]. In summary, this approach used an exact calculation of the estimated thermodynamics and expanded on prior work from the Mathews lab that calculated loop probabilities [151]. Among the predicted CDEs, known CDEs in the 3'UTR of *TNF*, *Ox40* and *ICOS* were correctly identified. Estimate folding probabilities of these known CDEs range between 15-98%. CDEII of *UCP3* is in the middle with 55%, while CDEI, which consists entirely of A-U base pairs, is significantly lower with 6% folding probability. Since *UCP3* CDEI is the experimentally proven CDE with the lowest folding probability [113], we used an estimate folding probability cutoff of 5%. CDEs with an even lower folding probability ($\leq$5%) may not be targeted by Roquin. About a third of the predicted CDEs had a folding probability $\leq$5%. In addition, we analyzed the evolutionary conservation of the predicted CDEs as an indicator for their functional importance. We assessed conservation by querying the predicted human CDEs in Multiz 46-way multiple sequence alignments with chimpanzee, mouse, dog and cow. CDEs were considered conserved if their nucleotide sequence was identical across all five species or if only tolerated nucleotide changes which retained CDE structure and corresponded to the initial descriptor (**Fig. 2.19**) were observed. Considered as tolerated were changes

in the triloop that still fit the definition of the proposed consensus (5' stem -YRN- stem 3'). It further included changes to the length of the stem, such that only 6 bp are required to be conserved and changes to the stem that can still form canonical base pairs, like compensating base pair changes, or G∘U wobble base pairs.

About every eighth CDE predicted in humans was found also in the other four species studied. Interestingly, the conserved CDEs had a significantly higher likelihood of folding (≥5%) compared to the non-conserved CDEs (**Fig. 2.21A**). The stem of CDEI in the *UCP3* wt element solely consists of A-U base pairs. Therefore, I analyzed the AU-content in the stems of all predicted CDEs. No significant enrichment for AU-rich stems could be observed for the conserved CDEs compared to the non-conserved CDEs (**Fig. 2.21A**). Strikingly, the conserved CDEs showed a significant enrichment of U at position three of the triloop compared to the non-conserved CDEs. At position one and two of the triloop, no significant change was observed between conserved vs. non-conserved CDEs. These findings are in accordance with the *in vivo* mutagenesis data, were triloops with a U at position three showed the best regulatory activity. Analysis

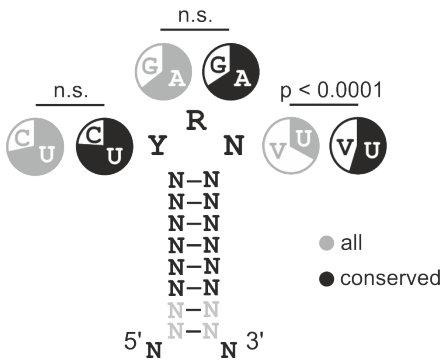

Figure 2.20: Nucleotide occurrence in the tri-nucleotide loop. Consensus used for bioinformatic prediction of new CDE structures by RNAMotif. Nucleotide occurrences in the triloop are shown for all and conserved elements. V = A, G, C. P-value < 0.0001 (Pearson's chi-square test). n.s. = not significant. Adopted and modified from [113].

of the data set also allowed me to evaluate the occurrence of certain CDE features that have previously been described as important for CDE based regulation. For example, while the mutation study on the *TNF* CDE indicated the need for a purine stack on the 3' side of the stem, the mutation study on the *UCP3* CDEs could not confirm this. Analyzing the occurrence of purine stacks, it can be seen that they only show a slight preference in the conserved CDEs compared to the non-conserved CDEs (**Fig. 2.21A**). Thus, I conclude that while purine stacks do occur in CDEs, they a not required for their activity *in vivo*. This confirms the results presented in this work. As described in section 2.3.2 the mutational analysis of the *UCP3* wt element showed that a G in position two of the loop was sufficient to significantly increase luciferase activity. This finding is in contrast to the *TNF* CDE were G and A are equally effective in regulating

mRNA half-life. A distinctive difference between the two CDE elements is the polarity of the closing base pair, which is R-Y (A-U) in both *UCP3* CDEs as opposed to Y-R (C-G) in the *TNF* CDE. Therefore, I analyzed if a G in position two in the loop occurs preferably in conjunction with a Y-R closing pair in conserved CDEs. With an A at position two in the loop no preference of the closing base pair was observed. In contrast, there was a significant enrichment for Y-R closing base pairs among G containing triloops in conserved CDEs compared to non-conserved CDEs (**Fig. 2.21B**). Taken together,

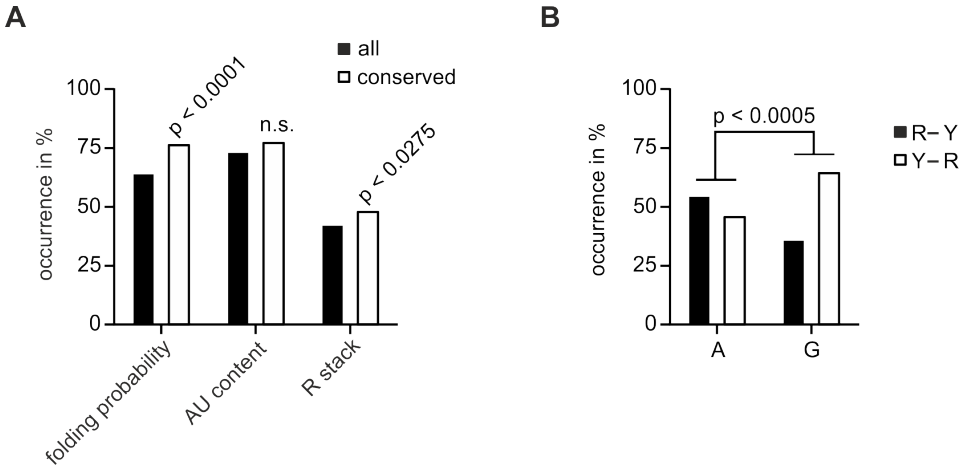

Figure 2.21: Statistically enriched CDE features. A) Probability of folding and composition of the stem of conserved CDEs. Comparison of folding probability ($> 5\%$), AU-content ($> 50\%$) and occurrences of purine (R) stacks at the 3' side of the stem (at least three consecutive purines) between conserved and non-conserved CDEs. p = p-value (Pearson's chi-square test). n.s. = not significant. **B)** Co-occurrence of A or G at position 2 in the triloop with R-Y or Y-R closing base pairs. R = A, G and Y = C, U. P-value < 0.0005 (Pearson's chi-square test). Adopted and modified from [113].

I showed that CDEs that do not correspond to the initially proposed consensus (**Fig. 3.1C**) are conserved across human, mouse, cow, pig and dog. These conserved CDEs are enriched for a higher estimate probability of folding and do show a strong preference for U at position three of the triloop. In addition to this, mutational analysis and bioinformatic prediction suggest that the purine in position two in conjunction with the polarity of the closing base pair fine tunes the repressive strength of a CDE.

2.4.2 *In vivo* Verification of New Roquin Target Genes

With the information obtained, I was able to compile a set of 166 high potential target genes that contained 1-2 CDEs in their 3'UTR. To be considered, the CDE had to be evolutionarily conserved, a folding probability $> 5\%$ and a U at position three of the triloop. A list of these high potential candidates can be found in **Suppl. Tab. 5.4**. To verify if the CDEs are functional, I chose three single and one tandem CDE to

analyze their influence on gene expression. Therefore, I fused the single CDEs of *CSF3* (also G-CSF), *CYR61* (also CCN1) and *ZC3H12A* (also Regnase-1 or MCPIP-1) and the tandem CDE of *TM2D3* together with 40 nt upstream and downstream endogenous sequences to the firefly luciferase. In order to attribute any effects directly to the CDEs, additional constructs were cloned in which point mutations were inserted that prevented triloop formation and thus no longer allow CDE-based regulation. An overview of the CDEs and the corresponding mutations is depiced in **Fig. 2.22**. As positive control for single CDEs the *TNF* CDE_{37} [56] and a corresponding loss of function mutant were fused to the Firefly luciferase. The *UCP3* wt element and *UCP3* CDEI/IImut (**Fig. 2.14**) were used as control for the *TM2D3* tandem CDE. In order to quantify the effect

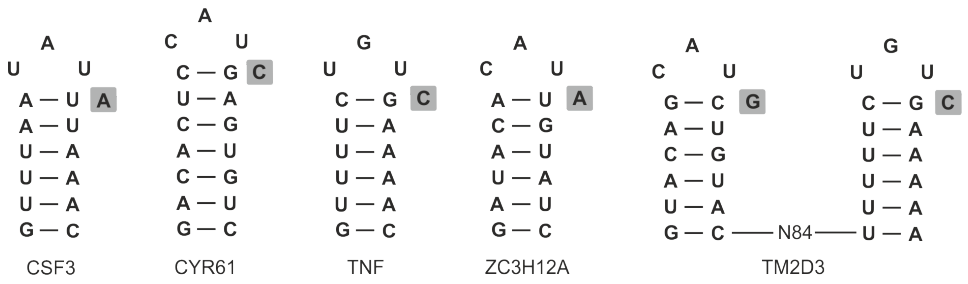

Figure 2.22: Overview of conserved CDEs in the 3'UTRs of new Roquin targets. CDEs were predicted by RNAmotif and selected based on their structural conservation, folding probability and a U at position three of the triloop. Inactivation mutations preventing triloop formation are shown next to the respective CDE. Adopted and modified from [113].

of the new CDEs on reporter gene activity, I measured luciferase activity in HEK293 cells. The single CDE reporter constructs reduced luciferase activity to 67% (*CSF3*) - 82%(*CYR61*) compared to the empty vector control (**Fig. 2.23**). These values are comparable to the individual CDEs from the *UCP3* wt element, where CDEI reduces luciferase activity to 73% and CDEII to 67% (**Fig. 2.14B**). In the luciferase reporter system the *TNF* CDE reduce luciferase activity to 63% compared to the empty vector control (**Fig. 2.23**). The tandem CDE from the 3'UTR of *TM2D3* reduced luciferae activity even further to 47% of the empty vector control which is comparable to the repression by the *UCP3* wt element. Most importantly, in all new CDEs studied, the mutation of the triloop to a pentaloop resulted in a significant increase in luciferase activity. This clearly demonstrated that the observed repression of luciferase activity is directly related to the presence of the respective CDE. In addition this proofed, that the new single and tandem CDEs were accurately detected by our computational approach.

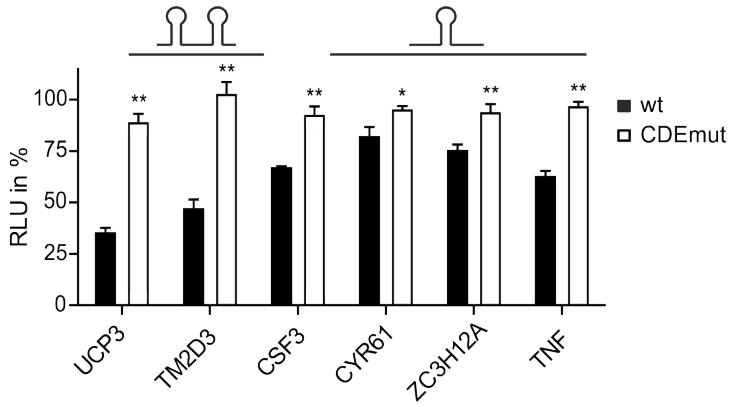

Figure 2.23: *In vivo* verification of new CDE-like elements. Influence of predicted CDE-like elements and mutants on luciferase activity in HEK293 cells. Firefly luciferase activity was normalized to *Renilla* luciferase as internal transfection control. Values are normalized to an empty vector control n = 3. (**) P-value < 0.01. (*) P-value < 0.05. Adopted and modified from [113].

2.4.3 Knockdown of Roquin Proteins Increases mRNA Abundance of CDE-encoding mRNAs

Next, I wanted to verify these finding in an endogenous context. Therefore, I quantified mRNA levels of CDE-encoding genes after knockdown of Roquin proteins. For my analysis, I used the four CDE-containing genes, from section 2.4.2. In addition to this, I also chose eight other CDE-encoding genes from the high confidence set (**Suppl. Tab. 5.4**). Because some of the predicted targets are known to alter the function of endothelial cells, I performed this analysis in HEK293 cells and HUVECs. For the knockdown of Roquin proteins in HEK293 cells, I reverse transfected HEK293 cells with siRNAs targeting human Roquin-1 and Roquin-2. 48 h post transfection RNA and protein samples were prepared. HUVECs were forward transfected 24 h post seeding. Likewise protein and RNA samples were prepared 48 h post transfection. Efficiency of the knockdown was analyzed by western blot and quantification of Roquin-1 and Roquin-2 mRNA levels (**Fig. 2.24**) to allow a direct comparison of relative protein levels and knockdown efficiency in HEK293 cells and HUVECs.

All twelve genes that contain a predicted CDE in their 3'UTR could be verified as Roquin target in HUVEC. mRNA levels of all genes significantly increased upon kockdown of Roquin proteins. This is in accordance with an increase of mRNA half-live in absence of Roquin proteins. The results substantiate that the bioinformatic prediction is very precise in the prediction of CDE encoding genes and therefore genuine Roquin targets. Of the selected genes, only ten were detectable in HEK293 cells. Interestingly, only for seven out of the ten genes an increase in mRNA level could be detected (**Fig. 2.25**). This suggests that not only the expression of Roquin target is cell type specific, the repression of target genes by Roquin proteins might be cell type specific as well.

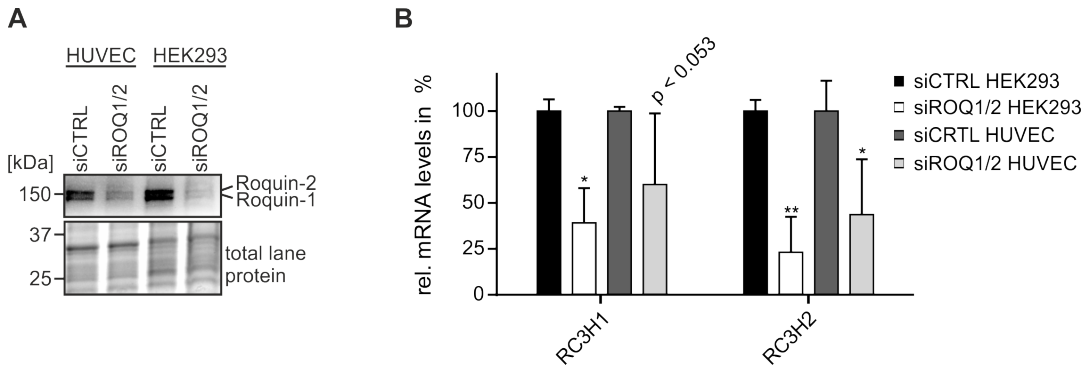

Figure 2.24: Knockdown of Roquin proteins in HEK293 cells and HUVECs. (A) Western blot of Roquin-1 and Roquin-2 after siRNA-mediated knockdown in HEK293 cells and HUVECs. Anti-Roquin was used to verify the respective knockdown. Total lane protein is shown as loading control. n = 3. **(B)** siRNA efficiency in HEK293 cells and HUVECs. RT-qPCR quantification of *RC3H1* and *RC3H2* mRNA levels after siRNA-mediated knockdown of Roquin-1 and Roquin-2 in HEK293 cells and HUVECs. Values are normalized to the housekeeping gene *RPLP0*. n = 4. (**) P-value <0.01. (*) P-value <0.05. n.d. = not detected. Adopted and modified from [113].

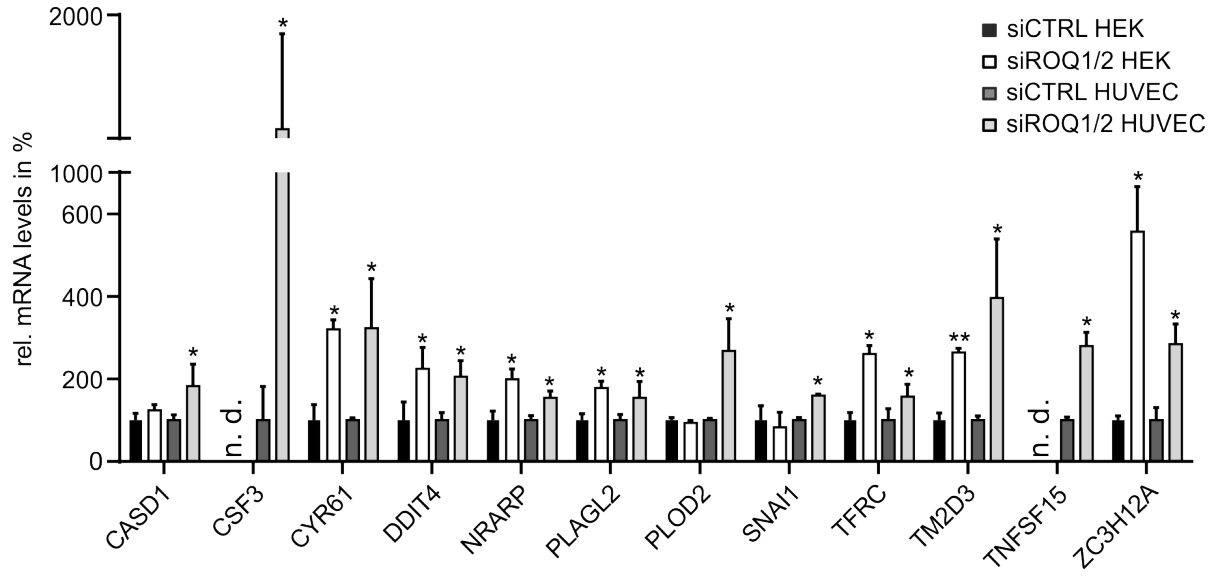

Figure 2.25: Target gene mRNA levels after Roquin-1/-2 knockdown. RT-qPCR quantification of new Roquin target genes encoding CDEs after siRNA-mediated knockdown of Roquin-1 and Roquin-2 in HEK293 cells and HUVECs. Values are normalized to the housekeeping gene *RPLP0*. n = 4. (**) P-value <0.01. (*) P-value <0.05. n.d. = not detected. Adopted and modified from [113].

2.4.4 Computational Prediction of ADEs

Recently, Janowski *et al.* discovered a new class of Roquin binding elements by SELEX. This alternative decay element (ADE) constitutes a 7 bp long stem capped by a U-rich hexaloop. In their study, they identified one endogenous ADE in the 3'UTR of *OX40*, which also harbors a canonical CDE, and demonstrated that it is bound by Roquin with high affinity [64]. No further ADEs have been identified since. Using our established bioinformatic pipeline, we scanned all human 3'UTRs for additional ADEs to gain insight how widespread this sort of Roquin binding site is on a genome-wide scale. For the descriptor search we used an ADE consensus that consists of a 6-8 bp long stem capped with a -GUUYUA- hexa-nucleotide loop (**Fig. 2.26A**). I chose the stem in analogy to functional CDEs. Similar to Roquin binding to CDEs, the structure of the Roquin-*Ox40* ADE-complex showed no sequence specific binding within the stem region [64]. By scanning all human 3'UTRs, 19 genes that encode for a potential ADE were

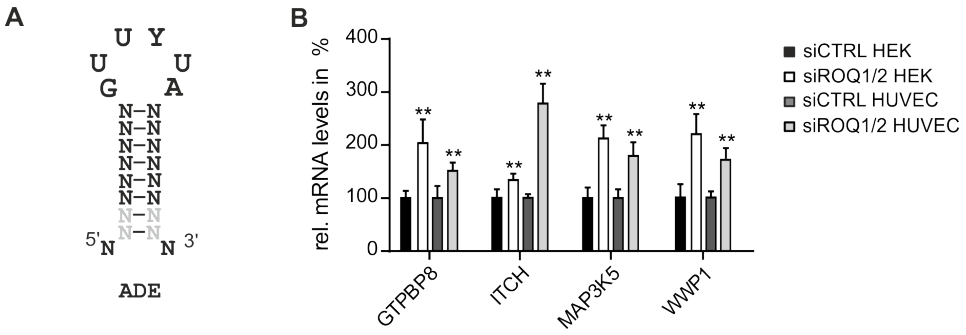

Figure 2.26: ADEs are independent Roquin binding sites. A) Consensus used for bioinformatic prediction of new ADE structures by RNAMotif. **B)** RT-qPCR quantification of new Roquin target genes encoding ADEs after siRNA-mediated knockdown of Roquin-1 and Roquin-2 in HEK293 cells and HUVECs. Values are normalized to the housekeeping gene *RPLP0*. n = 4. (**) P-value <0.01. Adopted and modified from [113].

predicted. Of these 19 only one ADE had an estimate folding probability < 5%. Similar to the CDEs, about 50% of the predicted ADEs had an AU-rich stem and a purine stack on the 3' side of the stem. In addition, no specific polarity for the closing base pair was observed in the putative ADEs. Of the 19 genes that contained a predicted ADE, three genes also encode for a canonical CDE. Interestingly, none of the three CDEs is evolutionarily conserved, suggesting that ADEs function independently as Roquin binding sites. Of the 19 predicted ADEs, six are conserved among chimpanzee, mouse, dog and cow. As demonstrated before conservation is a strong indicator for functional importance. In consequence, these six ADEs might be genuine Roquin targets. To test whether the ADE containing genes are regulated by Roquin, I analyzed mRNA abundance of four of the genes predicted to contain an ADE after knockdown of Roquin proteins in HEK293 cells and HUVEC. All four genes tested for regulation by Roquin

showed an significant increase in mRNA abundance upon Roquin knockdown in HEK293 cells as well as HUVECs (**Fig. 2.26B**). Most importantly, non of the selected genes contained an additional CDE, which highlights the self-sufficiency of ADEs as Roquin binding elements. In summary, our bioinformatic pipeline has enabled us to precisely identify new ADE containing genes and these genes are directly regulated by Roquin independent of the presence of a canonical CDE. This demonstrated that ADEs are indeed self-sufficient *cis*-regulatory elements by themselves.

In summary, redefining the consensus CDE and accurate genome-wide prediction of CDE containing genes as well as the genome-wide identification of ADE containing genes, enabled me to identify and verify a variety of *de novo* Roquin target genes. Many of the newly predicted and verified targets are not associated with the immune response, the proposed primary function of Roquin. Therefore, the results presented in this thesis suggest that Roquin proteins are involved in a much broader spectrum of cellular processes. These new roles for Roquin proteins provide great potential for deciphering post-transcriptional gene regulation and therefore need to be investigated in the future.

3 Discussion

In my doctoral thesis, I discovered a new tandem CDE and exploited it to discover previously undetected Roquin binding sites on a genome-wide scale, which allowed me to expand Roquins role in health and disease. For that, I first used a bioinformatic prediction to screen for structurally conserved regulatory RNA elements. With this method, I successfully identified regulatory RNA elements in 3'UTRs. The regulatory element in the 3'UTR of *UCP3* folds into a tandem CDE structure and is bound by Roquin proteins with high affinity. CDEs have previously been identified in a variety of immune response related mRNAs. In the context of this work, a Roquin target that is not connected to the immune response is being analyzed in depth, for the first time. Previously CDE-containing mRNAs were identified by a descriptor-based motif search based on the *TNF* CDE consensus. In my thesis, I revised the CDE consensus based on an in-depth mutational analysis of the *UCP3* tandem CDE. This allowed the prediction and identification of many novel decay elements in 3'UTRs of mRNAs that are connected to a variety of physiological processes. It demonstrates the importance of the Roquin-CDE interaction and its contribution to the regulation of critical cellular processes by Roquin-mediated repression [113]. With my work, I hope to initiate the further investigation of structurally conserved RNA elements in higher eukaryotes and decipher their contribution to post-transcriptional regulation of gene expression.

3.1 Identification of Regulatory RNA Elements Based on Structural Conservation

In the following, I will first present the starting point of my study to highlight the importance of researching regulatory active RNA elements. RNAs are the central hubs for regulating gene expression and protein synthesis [152] [153]. It is widely accepted, that its functionality does not rely on sequence content only, but is also directly attributable to its ability to fold into specific intricate secondary and tertiary structures [154] [131]. For the understanding of post-transcriptional gene regulation, it is therefore necessary to have precise knowledge of structure and location of these RNA elements within an mRNA of interest. There are many ways how RNA structure can be assessed experimentally. High-resolution structures can be obtained by X-ray crystallography [155] or nuclear magnetic resonance [156]. However, these methods have shortcomings like cost or manual labor, rendering them slow and low-throughput [157]. In addition, these methods only provide structural insights into already known RNA elements. On the other

hand structural profiling recently is emerging as an affordable and high-throughput approach to structure analysis as it provides snapshots of structural states at nucleotide resolution *in vivo* or *in vitro* [158]. It can be utilized to analyze an RNA structurome of a cell under different conditions and help to shed light on the role of structures in governing biological functions. Since all *in vivo* structural probing methods are based on cell permeable chemicals with different reactivity for single- and double-stranded regions [159] [111], they all share the principal drawback that it is unclear whether an undetected region, i.e. one that is not accessible to the chemical, is actually double-stranded or bound by a protein. This is of critical importance since it is known that 3'UTRs are extensively bound by a variety RBPs at the same time [160] [161]. Complementary methods circumvent this issue by detecting unbound double-stranded regions with psoralen-induced cross-linking or protein-bound regions by UV-induced cross-linking in combination with immunoprecipitation of the desired protein [112] [101]. Despite the fact that all these methods are promising tools for a comprehensive elucidation of the RNA structurome in the future, they share one significant disadvantage: they are limited to the transcripts present at the time of the experiment. Therefore these methods are not applicable for the detection of transcripts that are induced by stress or show cell type specific expression.

3.1.1 Prediction of Structurally Conserved RNA Elements

An alternative approach to detect specific RNA structures in cells is RNA secondary structure prediction. The key benefit of this approach is the possibility of identifying a regulatory element independent of the cellular level of the mRNA it is located in [114] [162]. Similar to sequence conservation, evolutionary conservation of an RNA structure can also be an indicator of functional importance [118]. A critical aspect for identifying a conserved structure is to exclude sequence conservation bias. Programs like Dynalign and Foldalign can identify conserved RNA structures by taking two unaligned sequences and simultaneously folding and aligning them to identify the lowest free energy structure both sequences can adopt [124] [125], whereas programs like RNAz [123] or Evofold [163] rely on a fixed sequence alignment as input. Therefore, Dynalign includes the advantage of phylogenetic conservation for structure prediction, without the disadvantage of relying on high sequence conservation. A disadvantage of the Dynalign prediction, which must be mentioned is that the algorithm is only able to compare two sequences with each other. In this thesis, a prediction based on all human and mouse UTRs was used, since at the time the analysis was performed those were the best annotated genomes. By only comparing two species, which are very similar, it is not possible to asses whether a structure really is conserved on a broader phylogenetic scale, which in turn increases the possibility of false positive predictions. In general genome diversity is low among mammals. The 3'UTRs of human and mouse genome have $\sim$73% sequence identity [164] [165]. Due to the high sequence identity, it can be assumed that the RNA secondary structure between both organisms will also be very similar, resulting in a large number

of predicted conserved structures. Including more genomes into the analysis should therefore increase the accuracy of the prediction of conserved regulatory structures. In order to achieve this, the ∼6000 initially predicted elements could be re-evaluated with the program Multilign. Multilign is based on multiple Dynalign calculations and can predict structures common to three or more sequences [166]. Reinspection of the windows with Multilign will considerably reduce the false positive rate, as it allows the comparison of the structural conservation in a broader phylogenetic context.

3.1.2 Screening for Functional RNA Elements

The three dimensional structure of RNAs can be essential for their respective physiological function. By predicting conserved structures in UTRs using Dynalign, I was able to identify functional structured RNA elements. The identification of regulatory RNA structures in higher eukaryotes remains a challenge, but is key to understand complex patterns in RNA regulatory networks. Here, I present how I investigated the regulatory properties of computationally predicted, structurally conserved, regulatory RNA elements. In order to asses the regulatory properties of the predicted elements, I performed a dual luciferase assay based screening to investigate the influence of 20 selected 3'UTR elements on luciferase activity. While the majority of predicted elements did not have an effect on reporter gene activity, four elements significantly reduced luciferase activity to ∼ 30-67% compared to the empty vector control (**Fig. 2.1**). For the experiments cells were maintained in a steady state under optimal growth conditions. However post-transcriptional gene regulation often occurs in response to stress signals [167] [168]. Therefore, modification of culturing conditions, i.e. hypoxia, iron depletion or serum starvation, could be applied to identify regulatory elements that exert their function under certain conditions. Such an approach would take full advantage of the unbiased predictive approach by Dynalign. Still, the use of a luciferase assay based screening system limits the number of elements that can be analyzed, because it cannot be up-scaled to a high-throughout methodology. In order to facilitate an efficient *in vivo* interrogation of RNA elements for their function in post-transcriptional gene regulation fluorescence-based screens have been used in a variety of contexts [169] [170]. These studies utilize fluorescence-based reporter libraries, which are analyzed and sorted by fluorescence activated cell sorting (FACS). In a study by Wissink *et al.* [107] they exploit chromosomally integrated dual-fluorophore reporter libraries, in combination with FACS, in order to sort cells into different pools of reporter gene activity. In order to recover the regulatory sequences, high-throughput sequencing is applied to the different pools to attribute regulatory impact to discrete sequences [107]. A similar strategy could be implemented with the plasmid pDF-FRT constructed in this thesis to query all predicted structurally conserved elements. This encodes mCherry and *GFP*. Both reporter genes are expressed from minimal promoters and share the same enhancer element. A multiple cloning site downstream of the GFP ORF allows to insert sequences of interest to investigate their influence on *GFP* expression. In addition, the plamid en-

codes an FRT site for genomic integration into a "Flip-In" host cell line. For maximum applicability, a pDF-FRT plasmid library that contains all 100 nt long windows would have to be cloned. Such a library could be easily integrated into different cell lines as well as analyzed under diverse cultivation conditions and stimuli. Such an experimental approach can also be combined with a knockdown of RBPs. In previous studies knockdown of Dicer was applied to identify mRNAs that are targets of RNA silencing [171]. Performing a Dicer knockdown on an integrated dual-fluorophore reporter library would therefore allow the identification of elements targeted by miRs. This approach can also be applied to RBPs that affect mRNA stability or translation efficiency. Similar studies were performed with the ARE binding proteins AUF1 [172] and TTP [173] or RBPs that recognize specific RNA structures, like Roquin [52], but these studies are always limited to one cell type and growth condition. Combining the presented computational prediction with an integrated fluorophore reporter library, combined with RPB knockdown and coupled to RNA sequencing may therefore be the most efficient approach for the *de novo* identification of regulatory RNA structures.

3.1.3 A Repressive RNA Element in the 3'UTR of *UCP3*

I discovered a repressive element in the 3'UTR of the *UCP3* gene. To proof the self sufficiency of the element, I initially analyzed different variants of the 1131 nt long *UCP3* 3'UTR and their potential to regulate a luciferase reporter (**Fig. 2.2**). Only reporter RNA encoding the predicted 100 nt structurally conserved RNA element reduced luciferase expression significantly. The residual 1031 nt long 3'UTR variant lacking the predicted element only exhibited a slight reduction of reporter gene activity [113]. This led me to the conclusion that the element, predicted by Dynalign, is responsible for the observed reduction of luciferase activity. Further analysis showed, that duplication of the element led to an increase in repression of the same magnitude. This supported my conclusion that the element harbored a binding site for a *trans*-acting factor. UCP3 is a member of a five mitochondrial membrane protein spanning subfamily (UCP1-5) of the SLC25 family of mitochondrial anion carrier proteins [136]. In general UCPs are important in regulating cellular energy metabolism and as attenuators of reactive oxygen species [147]. UCP3 is encoded in the nucleus and imported into mitochondria after translation of the entire protein in the cytoplasm and inserted into the inner membrane by an import machinery specialized for membrane proteins [174]. The most prominent member of the uncoupling protein family is UCP1 which is primarily expressed in brown adipose tissue. Extensive studies highlighted its thermogenic function in the context of oxidative phosphorylation [136]. There, its physiological function is to mediate a regulated, thermogenic proton leak [129]. The inducible conductance of protons across the mitochondrial inner membrane (MIM) represents a common property of all UCPs [175]. Importantly, the dissipation of protons across the MIM accounts for a substantial part of the resting metabolic rate [129]. Therefore UCPs represent potential targets for the treatment of metabolic diseases. While the involvement UCP3 in the pathogenesis of

cardiovascular diseases is widely acknowledged, its physiological role is still under debate [176] [128] [177]. Undoubted is its highly selective expression in skeletal muscle, brown fat and diabetic heart. Previous studies showed that although UCP3 has uncoupling properties, this does not seem to be its primary function. In a physiological context an increase in UCP3 protein does not necessarily increase uncoupling. This is supported by studies with $UCP3^{-/-}$ mice that did not show impaired thermoregulation, which suggests other cellular functions for UCP3 [178] [179]. Three main hypothesis are currently under debate: 1) UCP3 may play key regulatory roles in mitochondrial fatty acid oxidation; 2) UCP3 may protect cells against oxidative damage by mitigating ROS emission from the electron transport chain; 3) protect mitochondria from from oxidative damage by exporting lipid hydroperoxide (LOOH). In physiological studies elevated UCP3 levels were observed in obesity-resistant mice [180] and a decrease in UCP3 levels was associated with insulin sensitivity, a condition that precedes diabetes [181]. Thus, modulation of UCP3 levels is an interesting target for the development of novel therapeutics. Fasting, acute exercise and high-fat intake all lead to increased UCP3 levels, which is in line with the induction of UCP3 caused by PPAR transcription factors [147]. Interestingly, *UCP3* mRNA levels are elevated in fasted mice and are quickly restored after refeeding [182]. This indicates that the *UCP3* mRNA may be unstable. Thus, it can be concluded that a rapid constitutive decay of *UCP3* mRNA could be essential to trigger a rapid response upon changing nutritional states.

3.1.4 The UCP3 Repressive Element is not a miR Binding Site

Almost 60% of all protein-coding genes contain conserved miR target sites, predominantly in their 3'UTR [183] [36]. I initially tested whether the element serves as a single stranded miR binding site and thereby represses reporter gene activity. None of the miRs identified by TargetScan analysis are expressed in significant amounts in HEK293 or HeLa cells. Therefore, I concluded that repression is most likely not caused by miRs. In order to analyze if the element represents a miR target after all, I performed overexpression of miR-152 in the presence of luciferase reporters containing either the 100 nt long element or a mutant without the binding region for miR-152. The *UCP3* repressive element contains a highly conserved miR-152 target site and miR-152 is known to destabilize the mRNAs of many tumor related genes [184]. In addition, miR-152 serum levels positively correlate with T2DM [141]. In this experimental set up, luciferase activity was reduced to 70% in the wild type, but not in the miR binding site deletion mutant. These findings strongly argue against the involvement of miR-152 in repression by the *UCP3* element because repression by miR-152 is independent of the observed repression by the 100 nt element.

3.2 The Tandem CDE in the *UCP3* 3'UTR

To assess the features important for repression, I performed a truncation study. As expected deletion of the 5' end of the element, which contains the predicted miR target sites did not influence reporter gene repression. Further truncation from either the 5' or 3' end resulted in a significant loss of function. The residual 64 nt minimal motif contains two RNA hairpins that are similar in sequence and structure to CDEs. CDEs are recognized by Roquin proteins in a specific shape-dependent manner and repress target mRNAs by rapid mRNA decay [56]. To substantiate the assumption that the *UCP3* element adopts a CDE-like structure, I verified the structure of the *UCP3* element *in vitro* by in-line probing and and performed a mutational analysis that connected reporter gene repression to CDEI and CDEII. Interestingly, only CDEII could be observed by in-line probing of the 100 nt window (**Suppl. Fig. 5.6**), whereas the remaining sequence appeared single stranded. In comparison, probing of the 64 nt long minimal motif verified the existence of both CDEs. This finding suggests that the tandem CDE my sample different conformations in its natural sequence context. This was also observed with the *TNF* CDE in a similar manner. In-line probing analysis of the *TNF* CDE in a 150 nt sequence context resulted in diffuse cleavage patterns, compared to the *TNF* CDE_{37}. In addition, the stem-loop in CDE_{150} provided a much smaller decrease in free energy compared to the CDE_{37}, indicating that CDE_{150} may fluctuate between alternative or unstructured conformations. [185].

Reporter gene repression by the *UCP3* tandem CDE is caused by a rapid decrease of mRNA half-life. Accelerated mRNA decay is known to be caused by recruitment of the deadenylase machinery by Roquin to the CDE-encoding mRNA [56]. The observed mRNA half-life of 1.5±0.4 h for a GFP reporter gene encoding the *UCP3* tandem CDE is comparable to the mRNA half-lifes of other CDE-encoding genes like *NFKBIZ* 1.6±0.2 h, *LER* 1.7 ± 0.4 h, *PPP1R10* 2.1 ± 0.3 h and *HMGXB3* 1.6 ± 0.1 h [56]. RNA affinity purification with the *UCP3* tandem CDE and western blot analysis confirmed Roquin binding to both CDEs present in the *UCP3* element. EMSA experiments showed, that Roquin binds to both *UCP3* CDEs directly via its core ROQ domain which is consistent with previous reports [145] [56] [59]. The core ROQ domain of Roquin-1 binds with an apparent K_D of 138 nM to the UCP3 tandem CDE [113]. This is consistent with other high affinity Roquin binding sites that have been previously analyzed by EMSAs such as the *TNF* CDE with a K_D of ∼119 nM [56] [60] [59] and the *ICOS* CDE with a K_D of ∼136 nM [186]. In contrast, to the *TNF* CDE, the *UCP3* wt element needs two CDEs in close proximity to achieve such high affinity as mutation of either CDEI or CDEII significantly lowers affinity. Still the ROQ domain of Roquin-1 binds CDEmutII with similar affinity as the *Ox40* CDE with an apparent $K_D > 1000$ nM [64]. For the ROQ domain of Roquin-2 a K_D of ∼360 nM to the *Ier3* CDE has previously been reported [63], which is comparable to the affinity determined for the Rquin-1 ROQ domain interaction with CDEImut. In conclusion the apparent binding constants of the *UCP3* wt RNA and the single mutants are in the range of previously determined binding constants of Roquin

to CDEs. Interestingly, both CDEs are necessary for efficient Roquin binding and robust suppression of gene repression. Similar characteristics have recently been described for a tandem element from the *ICOS* 3'UTR, encoding a CDE and an octaloop-containing stem loop, by EMSAs [58]. Similar to my observations on the *UCP3* tandem CDE, efficient binding of Roquin N-terminus to the 88 nt long *ICOS* tandem element was only observed if both stem loops were present.

The *ICOS* 3'UTR harbors a similar composite Roquin binding site. It also contains two closely spaced Roquin binding sites. As in the *UCP3* 3'UTR the presence of both CDEs is necessary for efficient Roquin binding and subsequent *ICOS* repression [58]. In addition to that Roquin mediated repression by the *ICOS* 3'UTR is supported by a much further 5' localized Roquin-bound CDE-like stem-loop that has previously been identified [187] [53]. The data from the study by Rehage *et al.* indicate a cooperative function between the different Roquin binding sites in the *ICOS* 3'UTR. Interestingly, shortening the distance between the individual Roquin binding sites in the *ICOS* 3'UTR leads to increased regulation by Roquin. It is therefore possible that several Roquin binding sites lead to a positive cooperation and thus to a better regulation by Roquin. This hypothesis is also supported by my findings that *in vitro* high affinity binding of Roquin is only observed with the *UCP3* wt RNA where both CDEs are present. However to this end it is still unclear how such cooperativity is achieved at a molecular level. In solution Roquin protein is a monomer, and direct interaction of Roquin proteins has not been observed so far. However, interaction of Roquin proteins could be induced by RNA binding or by auxiliary factors. Just recently NUFIP2 protein has been discovered to enhance Roquin binding to *ICOS* mRNA *in vitro* and therefore enhance repression by Roquin *in vivo* [58].

Conspicuously, although two CDEs are present in the *UCP3* wt RNA only one RNA-protein complex band was observed in EMSAs with the core ROQ domain and the *UCP3* wt RNA. However, under equal running conditions, this band migrates significantly slower than the single bands observed for the two mutants, which only contain one CDE. Such underrepresentation of intermediates is a feature of cooperative binding. Cooperative binding can also be deduced from the sigmoidal binding curve of the core ROQ domain to the *UCP3* wt RNA. To unequivocally confirm that more than one core ROQ domain is bound to the *UCP3* wt RNA, I performed stoichiometric binding experiments. The analysis confirmed that at least two core ROQ domains are bound. Interestingly, two bands with retarded migration were observed for binding of the N-terminal Roquin fragment to the *UCP3* wt RNA. With the CDEImut RNA only the faster migrating band could be observed, which indicates only one bound N-terminal fragment. For the CDEIImut RNA again a second, slower migrating band at higher protein concentrations was observed, suggesting that the CDEII pentaloop mutation does not completely inhibit the binding of a second N-terminal fragment. This could have reasons that are independent of the binding of the core ROQ domain. First, the N-terminal fragment also contains a CCCH ZnF domain not present in the core ROQ domain. CCCH ZnF are known RNA binding domains that can recognize linear

RNA sequence motifs [188] [189]. Second, binding of the stem region of CDEII by a dsRNA-binding "b-site" found in extended ROQ domains [62] and also present in the fragment I used. Importantly, this residual binding has no effect *in vivo* as the CDEII pentaloop mutation is as effective as other more destructive mutations of CDEII. The Hill coefficient for binding of the core ROQ domain to *UCP3* wt RNA is 3, indicating cooperative binding, while the Hill coefficient for the N-terminal fragment is 1, indicating no cooperativity. Thus, these two isolated protein fragments show different behavior in *UCP3* wt RNA binding. Ongoing experiments by Dr. Andreas Schlundt (Johann Wolfgang Goethe-Universität, Frankfurt) currently investigate this contradiction by a structural analysis based approach.

3.2.1 Structure-Sequence-Function Analysis of the Tandem CDE

In 2013, Leppek *et al.* published that Roquin binds to a short stem-loop motif in the 3'UTR of *TNF* and induces mRNA decay [56]. Mutational analysis and a genome wide descriptor search identified ∼50 CDE-encoding transcripts. CDEs in these target transcripts were conserved among mammals. Contradictory, in the same study, a large number of Roquin-1 targets were identified, which did not contain detectable CDEs. Based on RNA-IPs [56], PAR-Clip [52] and SELEX [64] approaches Roquin-1 binding to non CDE-like motifs and its contribution to mRNA destabilization was demonstrated in various studies. Still, the *TNF* CDE remains the only CDE for which a comprehensive *in vivo* mutational analysis has been performed and its functional consequences analyzed [56]. The findings of this initial study have been extended by several *in vitro* binding studies as well as structural biology approaches to unravel the requirements for a functional CDE. Based on these findings different CDE consensus have been described (**Fig. 3.1**). Both CDEs in the *UCP3* wt element do not correspond to the previously proposed CDE consensus. A precise mutational analysis revealed that Roquin binding is primarily shape-dependent. This is in accordance with previous structural analysis [57], but has not been demonstrated *in vivo* yet, where at minimal pyrimidine-purine base pairs in the apical part of the stem are required for Roquin binding activity. In this regard, the data presented in this thesis contradict previous findings. In particular, the necessity of pyrimidine-purine base pairs in the upper part of the stem was convincingly refuted by the mutation study of the *UCP3* tandem CDE. Even G∘U wobble base pairs remain a certain regulatory capacity. This leads to the conclusion that the nucleotide composition of the stem is not essential for *in vivo* activity. Mutational analysis of the tri-nucleotide loops of CDEI and CDEII largely confirms the previously suggested -YRN- motif. Contrary to *TNF* CDE, A in position two of the loop (instead of G) is required for efficient *in vivo* activity. Also the previously proposed necessity of a purine stack at the 3' side of the stem [56] [60] could not be recapitulated in any of the *UCP3* CDEs. In addition, I was able to demonstrate for the first time that a minimum stem length of 6 bp is required for a fully active CDE [113]. This complements previous findings that extending the stem to 9 bp inhibits CDE activity [56].

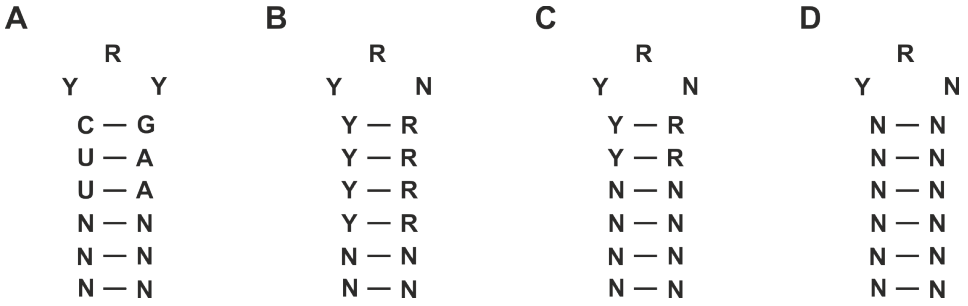

Figure 3.1: Evolution of the CDE consensus. Roquin recognizes a relaxed RNA stem-loop *in vivo*. **(A)** *TNF* CDE derived consensus by [56]. **(B)** *In vivo* CDE consensus based on mRNA decay assays in 21,54 and suggested by Codutti *et al.* [60]. **(C)** A general suggested CDE based on *in vivo* and *in vitro* studies proposed by Schlundt *et al.* [57]. **(D)** New, relaxed general CDE consensus based on the data presented in this study [113].

Summarizing the collected data on the *UCP3* tandem CDE allowed for the formulation of a new consensus for active CDEs (**Fig. 3.1D**). Compared to previously proposed CDE consensus (**Fig. 3.1A-C**), it is composed of a 6-8 bp long stem of variable sequence and a -YRN- tri-nucleotide loop.

3.2.2 Identification of New Roquin Target Genes

With the new consensus, I was able to predict and verify a large number of new Roquin targets and therefore expand Roquin's role in regulation of gene expression. As expected some of the identified target mRNAs are in line with the previously reported function of Roquin in immune response and inflammation. The data presented in this thesis clearly demonstrates that the mRNAs of the proinflammatory factors *CSF3*, *CYR61* and *TNFSF15* (also known as TL1A) are targets of Roquin mediated RNA decay. Most interestingly, the majority of new targets is associated to other cellular functions and therefore suggest a much broader involvement of Roquin in a variety of cellular processes. This is also consistent with the observed perinatal lethality of Roquin-1 or Roquin-2 knockout mice, which cannot be explained based on Roquin's function in the immune system.

Among the newly confirmed Roquin target mRNAs, I identified Regnase-1 as Roquin target. In addition, I successfully correlated the repression by Roquin with the predicted CDE. Interestingly Regnase-1 has been reported to detect similar RNA stem-loop structures in 3'UTRs of its target mRNAs [190]. Regnase-1 is a ZnF-containing RBP that also contains a PIN-like RNase domain and destabilizes its targets by endonucleolytic cleavage. Remarkably, Roquin and Regnase-1 share an overlapping set of mRNA targets. These target mRNAs can be suppressed in a cooperative manner [191] or by a spatiotemporal distinct mechanism [190]. One study demonstrated that both proteins function in

distinct subcellular locations in HeLa cells. Roquin is associated with processing-bodies and stress-granules, where it facilitates accelerated mRNA deadenylation and subsequent RNA degradation. In contrasts, Regnase-1 has been demonstrated to localize to the endoplasmic reticulum and active translating ribosomes, where it degrades its targets in a UPF1-dependent manner [190].

In accordance with the reported redundancy of Reganse-1 and Roquin in recognition of stem-loop structures, I identified mRNAs as Roquin targets that have previously been associated with Regnase-1, such as *TM2D3*, which contains a tandem CDE similar to the *UCP3* wt element, and *TFRC*. In a study by Mino *et. al.* *TM2D3* mRNA was significantly enriched by Regnase-1 immunoprecipitation [190]. Moreover, Yoshinaga *et al.* demonstrated that *TFRC* levels are significantly elevated in $Reg^{-/-}$ mice compared to control animals [192]. Additionally they performed luciferase based reporter assays that demonstrated that repression by Regnase-1 is mediated by an RNA stem-loop structure similar to CDEs independently of known iron responsive elements in the *TFRC* 3'UTR [192].

Cui *et al.* recently investigated cross-regulation between Roquin and Regnase-1. In contrast to the results of Jeltsch *et al.* in TH17 cells and my results in HEK293 cells and HUVECs, they found that the double knockout of Roquin-1 and Roquin-2 did not increase Regnase-1 mRNA levels in Jurkat cells [193]. Furthermore, overexpression of Roquin had no effect on the activity of a luciferase-Regnase-1 3'UTR-reporter in HeLa cells. Conversely, Regnase-1 knockout increased Roquin-1 and Roquin-2-mRNA levels in Jurkat cells. Also, overexpression of Regnase-1 reduced the activity of luciferase reporters with Roquin-1 and Roquin-2 3'UTRs in HeLa cells. Jeltsch *et al.* found elevated Regnase-1 protein levels after acute deletion of Roquin-1 and Roquin-2 in TH17 cells, leading to the conclusion that Regnase itself is a Roqion target [191]. Additionally and in accordance with our results, the 3'UTR of Regnase-1 encodes a functional CDE [194] similar to the 3'UTR of Roquin-1 and Roquin-2 [56]. Taken together, this allows for an extensive auto- and cross-regulation between all three proteins, but cross-regulation seems to be cell type specific [193] [191]. In conclusion this means, that Regnase-1 targets can either be regulated by Regnase-1 or Roquin directly or indirectly by Roquin mediated modulation of Regnase-1 mRNA abundance. From that, I propose that Roquin will share cellular functions with Regnase-1, additional to their function in the immune response, such as the regulation of the iron metabolism.

Interestingly, *in vivo* more CDEs seem to provide more efficient repression (**Fig. 2.2B** and **Fig. 2.23**). This does not necessarily require cooperative binding, but could also be due to Roquin's increased avidity to the UTR and thus more efficient recruitment of the mRNA decay machinery. Furthermore, this increased repressive activity of more CDEs only applies in isolation, i.e with short 3'UTR fragments as it is not transferred to the situation within a complete, endogenous 3'UTR. The highest induced mRNAs after Roquin knockdown are *ZC3H12A* (Regnase-1) in HEK293 cells and *CSF3* in HUVECs. Both encode only one CDE, while for example *CASD1* mRNA, encoding two CDEs, responds only slightly to Roquin knockdown (**Fig.2.25**). Thus, even if there is

cooperative binding of Roquin to RNAs *in vivo*, the sensitivity of an mRNA target to Roquin levels might depend much more on the overall composition of its 3'UTR (see also section **3.2.4**).

For effective binding of Roquin to the *UCP3* tandem CDE both CDEs are necessary as mutation of either CDE aggravates the affinity of Roquin to the *UCP3* wt RNA, which in turn reduces repression of gene expression *in vivo*. The 3'UTR of *TM2D3* also contains two CDEs in close proximity, similar to the *UCP3* 3'UTR. The two CDEs are seperated by 84 nt compared to 18 nt in the *UCP3* 3'UTR. Both tandem CDEs reduce luciferase reporter gene activity significantly stronger that any of the tested single CDEs. In contrast to the reporter gene based findings, the induction of mRNA levels after Roquin knockdown varies greatly between different target genes and also between cell types [113]. This leads to the conclusion that the level of repression correlates with the number of CDEs present in a UTR and therefore that two CDEs lead to a stronger repression of gene expression than one CDE. However, this somewhat digital behavior was only observable in isolation. This is nicely exemplified with the cytokine *CSF3* mRNA. Among all verified target genes it shows the greatest induction upon knockdown of Roquin proteins, yet the *CSF3* CDE alone reduces luciferase activity comparable to the other individual CDEs analyzed in this study.

In this context, it is must be mentioned that the *CSF3* CDE was identified as a repressive RNA element even before Roquin-mediated repression was discovered [195]. Putland *et al.* identified a stem-loop destabilizing element (SLDE) in the 3'UTR of *CSF3*. Even without knowing that repression is mediated through Roquin binding they found that the SLDE rapidly decreases mRNA half-life by increasing the rate of deadenylation [195]. Mutational analysis of the *CSF3* CDE yielded the same preference for a -YAU- tri-nucleotide loop motif. Additionally, they showed that triloop formation is mandatory for the rapid degradation of a reporter gene by the CDE. Similar to both *UCP3* CDEs the closing base pair of the *CSF3* CDE is A-U (R-Y) which further contradicts previous assumptions based on the *TNF* CDE that Y-R closing pairs are required for functional CDEs [56] [60] [59].

The mutational analysis data of the *CSF3* [195], *TNF* [56] and *UCP3* [113] CDEs led me to the conclusion that although -YRN- tri-nucleotide loops combined with any possible closing base pair are tolerated in active CDEs, some sequence combinations seem to be better recognized by Roquin and therefore are more efficient in regulating gene expression. The origin for such preferences is still unclear and has yet to be analyzed in detail. Being based on the initial findings on the *TNF* CDE all available high-resolution structures of Roquin-RNA complexes use a U-G-U tri-nucleotide loop with a C-G closing base pair [59] [61] [62] [63]. In those examples the G at position 2 of the loop can stack on top of the G of the C-G closing base pair and this interaction is locked by arginine 219 of Roquine-1 by a further stacking interaction [59]. The complex structure of the *TNF* CDE and Roquin-1 is depicted in **Fig. 3.2**.

An A at the central position of the loop in conjunction with R-Y closing pairs could allow for alternative interactions within the CDE or with the Roquin-1. These alternative

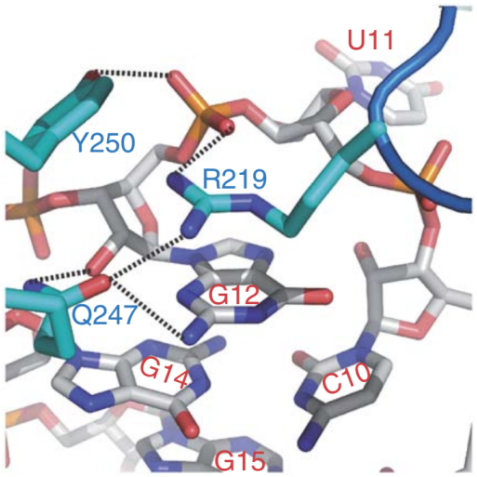

Figure 3.2: Structure of the Roquin-1 *TNF* CDE complex. Close-up view of the contacts between the ROQ domain and the G a position 2 of the *TNF* CDE tri-nucleotide loop. G at position 2 in the loop (G12) stacks with the G (G14) of the closing C-G base pair and is locked in that position by stacking interaction with Arg214. Figure adopted and modified from [59].

interactions could possibly support the formation of the CDE Roquin-1 complex and thus modulate *in vivo* CDE activity. Also, previous studies demonstrated that an incorrect preformation of the *TNF* CDE mutant lead to an inhibition of Roquin binding [60]. Therefore it could be possible that specific combinations of closing base pairs and loop nucleotides could impact preformation of the CDE and thereby facilitate or impede conformational selection of a CDE by Roquin.

3.2.3 ADEs are Self-sufficient Roquin Binding Sites

In Adittion to CDEs, Roquin binding to ADEs has been shown [64]. Based on the ADE from the *Ox40* UTR, the only example of an ADE so far, we predicted ADEs on a genome wide scale. Being constrained to the sequence of the hexaloop we only recovered a small number of 19 ADEs. Nevertheless, all further analyzed mRNAs encoding an ADE were induced upon Roquin knockdown. Strikingly, two of the verified targets, *MAP35K5* (ASK1) and *Itch* were previously described as Roquin targets, but do not contain a detectable CDE in their 3'UTR. MAP35K5 is a protein kinase involved in apoptosis signal regulation and is known to activate its targets c-Jun N-terminal kinase and p38 in diverse arrays of stresses [196]. Therefore the stability of MAP3K5 protein is regulated through ubiquitination by Roquin-2 [197]. My findings suggest that in addition to the regulation of proteolysis there exists an additional layer of *MAP35K5* regulation by targeting its mRNA stability. Essig *et al.* recently identified *Itch* as target of Roquin in T-cells [65]. In their study they showed that the 3' half of the *Itch* 3'UTR is sufficient for the repression of a reporter gene by Roquin proteins. In this region of the *Itch* 3'UTR

the predicted ADE is located. This further emphasizes that we accurately predicted functional ADEs and that they can serve as self-sufficient *cis*-regulatory elements.

3.2.4 Determinants of Roquin-mediated Regulation

The induction of CDE- and ADE-encoding target mRNAs upon knockdown of Roquin-1/-2 varies significantly between individual genes. In addition to that, some targets are sensitive to Roquin depletion in HUVECs but not in HEK293 cells (**Fig. 2.25**). This finding led me to the conclusion that in addition to the presence of a Roquin binding element further determinants are necessary to modulate the response of target genes to changes in Roquin levels. A prominent determinant of Roquin-mediated regulation is the folding probability of a given CDE/ADE in its respective 3'UTR sequence environment. High folding probability will facilitate recognition of the CDE/ADE by Roquin. This is also supported by my findings that the predicted conserved CDEs have a significant higher folding probability compared to the non-conserved CDEs. I addition to that, lower folding probability could be compensated in composite Roquin binding sites that harbor multiple CDE-like structures with sub-optimal properties that act cooperatively to repress mRNA levels *in vivo*. This is also supported by Essig *et al.* who showed that the minimal response element in the *Nfkbid* 3'UTR is composed of six stem–loop structures that cooperate to exert robust post-transcriptional regulation [54]. In addition, recent findings highlight the role of auxiliary sequences, such as stem loop elements with large U-rich loops, that are recognized by the ROQ domain [52] [58] [65], or AU-rich sequences that are recognized by the ZnF [52] that positively influence Roquin binding. Also, auxiliary proteins such as *NUFIP2* have recently been described to support the binding of the Roquin N-terminus to the *ICOS* and *OX40* binding elements [58].

Furthermore competing factors that also bind to CDEs will modulate the regulation by Roquin. As demonstrated by Cui *et al.* Regnase-1 could compensate diminished quantities of Roquin by destabilizing shared target mRNAs [193]. This is of particular importance as Regnase-1 itself is a target of Roquin and and therefore induced by low Roquin protein levels [113] [193] [191]. On the other hand, ARID5A [198] and BAG3 [199] have recently been identified as Roquin antagonists which increase the stability of target mRNAs.

In addition, the continuous destabilization of a target mRNA could be ensured by backup systems that function independent of Roquin-mediated repression. Independent mRNA destabilization could be achieved by AU-rich elements. A prominent example of Roquin independent repression is *TNF* which, in addition to a CDE, harbors an ARE in its 3'UTR that mediates mRNA destabilization by interaction with TTP [200]. It was also shown that Roquin can regulate target mRNAs in cooperation with miRs. Srivastava *et al.* showed that Roquin enhances Dicer-mediated processing of pre-miR-146a and directly interacts with Argonaute 2 and miR-146a, a miR that targets *ICOS* [201]. In addition, a study by Essig *et. al* indicates that binding of Roquin to a stem-loop structure precludes binding of miR-17∼92 and therefore stabilizes *Pten* mRNA [65].

The *UCP3* 3'UTR contains miR binding sites 5' to its tandem CDE [113]. I propose that Roquin therefore could also support post-transcriptional regulation of *UCP3* by these miRs.

Taken together, these findings provide an insight into the magnitude of Roquin-mediated regulation. A major challenge for the future will therefore be to fully understand how the regulation of individual target mRNAs by Roquin is achieved when different *trans*-acting factors and different *cis*-regulatory elements cooperate or antagonize each other to modify cellular functions.

3.3 Summary and Outlook

In this thesis, I redefined the binding preferences of Roquin and thereby expanded the role of CDE-mediated post-transcriptional regulation. My initial focus was the identification of regulatory RNA structures in mammals. For that, I have successfully implemented genome-wide bioinformatic predictions of conserved structured elements [132] for the identification of repressive 3'UTR RNA elements. This included a luciferase reporter based screening, which identified several regulatory elements. The *UCP3* 3'UTR element was most efficient in repression of reporter gene activity. I demonstrated that the minimal element able to exert full regulation is a structured RNA motif that folds into a tandem CDE structure. Structure formation of both CDEs is mandatory for function. RNA affinity purification enriched the paralogs Roquin-1 and Roquin-2. Roquin mediates rapid mRNA decay of *UCP3* mRNA. For efficient binding of Roquin to the *UCP3* 3'UTR both CDEs are required. In-depth mutational analysis redefined the *in vivo* requirements for Roquin binding. With the help of a descriptor based prediction of CDE-encoding genes, I was able to discover Roquin targets on a genome-wide scale. Combination of mutation analysis and bioinformatic prediction provided 160 vertebrate mRNAs with highly conserved CDEs and 19 ADE- encoding mRNAs. I provided evidence that mRNA stability of these transcripts is regulate by Roquin. Furthermore, I discovered that not only the expression of Roquin targets is cell type specific, also Roquin based regulation varies between different cells. Most importantly many of the identified targets are not related to the previously described role of Roquin in inflammation and the immune response. Therefore my data highlights additional, yet unstudied cellular functions of Roquin.

With the discovery of increasingly complex patterns of post-transcriptional gene regulation, it is becoming increasingly important to combine different genome-wide approaches with detailed mutational analysis *in vivo* and *in vitro* to understand the exact binding preferences of individual RBPs. A genome-wide prediction of RNA structures is therefore complementary as well as informative to *in vivo* experiments.

Future work should focus on improving the accuracy of the prediction of conserved RNA structures. Key to accurate structure prediction will be the availability of a large number of and better annotated mammalian genomes as well as increasing computational

resources. In the meantime, UTRs should be reevaluated for structural conservation with the algorithm Multilign. In addition, a high-throughput method for the analysis of putative regulatory element has to be established. I suggest, a dual fluorescent reporter gene library based on the pDF_FRT plasmid which could be integrated in the genome of different cell lines and allows for FACS analysis coupled to sequencing.

In addition, the increasingly complex binding of Roquin to composite RNA elements should be in the focus of future work. Therefore the question must be addressed, if Roquin binds to the tandem CDEs in a cooperative manner. NMR titration studies with the UCP3 tandem CDE and Roquin-1 protein may solve this question. Of particular importance will be the identification of auxiliary factors that modulate Roquin binding. Therefore, further studies need to investigate what exactly defines *in vivo* Roquin binding sites and identify the effect of auxiliary and competing proteins. Mutational analysis of new Roquin binding sites should be performed to extend our currently limited understanding of *in vivo* Roquin binding sites. For the dissection of Roquin-mediated protein-protein interactions co immunoprecipitation analyzed by mass spectrometry should be applied. This could be complemented by *in vivo* interactome studies of known Roquin targets. In total, this would allow the accurate prediction of target gene response to fluctuating Roquin levels and thus advance our understanding of the role of Roquin in health and disease.

4 Material and Methods

4.1 Materials

All materials, chemicals and technical equipment used in this study are listed in detail in the Appendix.

4.2 Methods

All commercially available kits have been used according to the manufacturer's recommendations unless otherwise stated.

4.2.1 Cell Culture and Transfection

Cell lines used in this study are listed in table 4.1

Table 4.1: Cell lines used in this study

Cell line	Description and ATCC number
C2C12	mouse (*Mus musculus*) near-tetraploid myoblast cell line (ACC 565)
HeLa HF1-3	human (*homo sapiens*) epithelial cervix adenocancinoma "Flp-In" Host Cell Line (ref)
HEK293	human near-triploid embryonic kidney cell line (ACC 305)
HUVEC	human umbilical vein endothelial cells (Lonza, CC-2519)

HEK293 cells and HeLa "Flp-in" Host Cell Line HF1-3 were cultured in Dulbecco modified eagle medium (DMEM, Sigma-Aldrich) supplemented with 10% fetal bovine serum (FBS, Sigma-Aldrich), 1 mM sodium pyruvate (Thermo Fischer Scientific) and penicillin-streptomycine (Pen-Strep, Thermo Fisher Scientific). HeLa HF1-3 cells were additionally supplemented with 100 µg/ml zeocine (Invivogen). HF1-3 cells stably expressing GFP-*UCP3* wt or -MUTI/II constructs were suplemented with 150 µg/ml hygromycine B (Invivogen). C2C12 murine myoblast cells were cultured as HEK293 cells but with 20% FBS (Sigma-Aldrich). For differentiation of C2C12 cells, the cells were grown to 90-100% confluency. Then cultivation medium was changed to Dulbecco modified eagle medium (DMEM, Sigma-Aldrich) supplemented with 2% horse serum (Sigma-Aldrich), 1 mM

sodium pyruvate (Thermo Fischer Scientific) and Pen-Strep (Thermo Fisher Scientific). HUVECs were cultured in EBM-Plus medium (Lonza) supplemented with EGM-Plus SingleQuots (Lonza) and 10% FBS (Thermo Fisher Scientific). All cells were maintained in T75 cell culture flasks at 37°C in a 5% CO_2 humidified incubator. When cells reached 70-90% confluency the culture medium was removed and cells were rinsed with 5 ml phosphate buffered saline (PBS, Sigma-Aldrich).Then cells were detached by incubation with Trypsin-EDTA (Invitrogen) for 5 min at 37°C. Trypsine digestion was quenched by adding 7 ml culture medium. The number of viable cells was determined by Trypan blue staining (Biorad) and counting the cells with a Biorad TC-10 automated cell counter. $0.5 - 1 \times 10^6$ HEK293 or HeLa cells or 2.5×10^4 C2C12 cells were transferred in a new culture with 20 ml culture medium. For the transient transfection of pDLP derived plasmids bearing the different *UCP3* variants Lipofectamine 2000 (Thermo Fischer Scientific) was used as recommended by the manufacturer. Therefore 1×10^5 HEK293 cells were seeded in 24-well cell culture dishes. A total amount of 100 ng of plasmid DNA per well was used for transfection. For the transfection 1 µl of Lipofectamine was diluted in 49 µl Opti-MEM (Sigma Aldrich) and incubated for 5 min at room temperature (RT). In the meantime the Plasmid DNA was diluted in 50 µl Opti-MEM. Both Dillutions were mixed in a 1:1 ration and incubated for 20 min at RT. The transfection mixture was than added to the the cells. 24 h post transfection a dual luciferase assay was performed. For miR-152 over expression, 400 ng pCMV-miR-152 or pCMV-MS and 100 ng pDLP-UCP3 reporter plasmid were cotransfected using Lipofectamine 2000 as described above. For the generation of RNA and protein samples the complete procedure was scaled up (2x) to 12-well cell culture dishes.

4.2.2 Plasmid Construction

Bacterial strains:

Table 4.2: Bacterial strains used for molecular cloning and plasmid production.

Strain	Genotype	Company
E. coli TOP10	F- mcrA Δ(mrr-hsdRMS-mcrBC) Φ80lacZΔM15 ΔlacX74 recA1 araD139 Δ(araleu)7697 galU galK rpsL (StrR) endA1 nupG	NEB

pDLP (luciferase reporter system):

To generate luciferase reporter plasmids encoding the 3'UTR of UCP3 (ENST00000314032.8), the 1131 bp long 3'UTR plus 98 nt endogenous context were PCR amplified using oligonucleotides UCP3_3UTR_fwd and UCP3_3UTR_rev and HEK293 genomic DNA

as a template. The PCR product was purified (QIAquick PCR Purification Kit, Qiagen) and digested using the restriction endonucleases NotI-HF (NEB) and HindIII-HF (NEB) as recommended by the manufacturer. After column purification (QIAquick PCR purification kit, QIAgen) the insert was introduced downstream of the firefly luciferase open reading frame into the multiple cloning site of pDLP [202] using NotI and HindIII restriction sites using T4 DNA Ligase (NEB). A 20 µl ligation reactions contained 25 ng linearized plasmid DNA and a 5-fold molecular excess of the insert DNA, 1 mM ATP, 2 µl T4 DNA Ligase buffer (10x, NEB) and 1 µl T4 DNA Ligase (NEB). Ligation was performed at 4°C over night. The 3'UTR deletion mutant (DEL) without the predicted structurally conserved window was generated by a two-step overlap extension PCR [203]. To generate the fragments for overlap PCR the oligonucleotide pairs UCP3_3UTR_fwd and UCP3_DEL_rev as well as UCP3_3UTR_rev and UCP3_DEL_fwd were used. Oligonucleotide sequences are listed in **Suppl. Tab. 5.7**. All 100 nt long variants of the UCP3 repressive element, other conserved CDEs or other Dynalign-predicted windows were generated by hybridization of complementary oligonucleotides and phosphorylation of restriction sites. E. coli TOP10 (NEB) were transformed with 10 µl of the ligation reaction and plated on lysogeny broth plates (LB; 1% tryptone, 0.5% yeast extratct, 1% NaCl, 2% agar) suplemented with ampicillin (100 µg/ml, Roth). Colony forming units (CFU) were analyzed for correct insert size by colony PCR. Positive tested CFUs were inoculated in 4 [ml LB medium with ampicillin and incubated az 37°C over night with agitation. Plasmid DNA was isolate using the QIAprep Spin Miniprep Kit (QIAgen) according to the manufacturers instructions. Mutations, cloning boundaries and coding sequences were verified by Sanger sequencing (Seqlab). All sequences are summarized in **Suppl. Tab. 5.2** and **Suppl. Tab5.5**.

pHDV (in vitro transcription):

For RNA synthesis, UCP3 wt, MUTI, MUTII, MUTI/II, R1, CDEImut, CDEIImut, CDEI/IImut sequences together with the T7 promoter (5'-TAATACGACTCACTAT AGG-3') at the 5' end and the HDV cleavage site (5'-GCCGG-3') at the 3' end, were generated by hybridization of complementary oligonucleotides and phosphorylation of complementary overhangs. The generated inserts were introduced into the EcoRI and NcoI sites the pHDV plamid [204] using T4 DNA ligase (NEB) as described above. In order to prepare plasmid DNA for *in vitro* transcription the QIAprep Spin Maxiprep Kit (QIAgen) was used according to the manufacturers instructions. RNAs were transcribed as HDV ribozyme fusions to obtain uniform 3' ends. Mutations, cloning boundaries and coding sequences were verified by Sanger sequencing (Seqlab). All sequences are summarized in **Suppl. Tab. 5.2**.

pCMV (miRNA expression):

For the overexpression of hsa-miR-152-5p, the genomic locus with 200 bp up- and downstream endogenous context was amplified using the oligonucleotides miR152_fwd and miR152_fwd (**Suppl. Tab. 5.2**). After purification the PCR product was digested using the restriction endonucleases XbaI (NEB) and XmaI (NEB). After purification of the digetsion reaction the insert was cloned in the XbaI and XmaI sites of the pCMV-MS plasmid [202] using T4 DNA Ligase (NEB) as above.

pDF_FRT (stable integration):

For generation of the integration plasmid pDF_FRT, eGFP and mCherry expressed from a shared CMV enhancer/promoter element were amplified using the oligonucleotides pDF_fwd and pDF_rev with pBI-CMV1 (Clontech) as a template. FRT-site, the hygromycin B resistance, a pUC origin and an ampicillin resistance were amplified using oligonucleotides pFRT_fwd and pFRT_ref (**Suppl. Tab. 5.2**) with pcDNA5/FRT plasmid (Invitrogen) as a template. Oligonucleotide sequences are listed in **Suppl. Tab. 5.7**. Both PCR products were digested using NotI-HF (NEB) and NheI-HF (NEB) and ligated using T4 DNA ligase, resulting in the pDF-FRT plasmid. Pasmid sequence was verified by Sanger sequencing (Seqlab).

pDF_FRT_UCP3_wt and _MUTI/II were generated by hybridization of complementary oligonucleotides and and phosphorylation of restriction sites. 3'UTR inserts were introduced downstream of eGFP ORF into the NotI and HindIII restriction sites of pDF_FRT using T4 DAN Ligase (NEB) as described.

4.2.3 Hybridization and Phosphorylation of Oligonucleotides

To generate inserts <120 nt, complementary oligonucleotides were annealed. The oligonucleotides were designed to form complementary overhangs for the direct use in ligation reactions with restricted plasmid DNA. Initially 50 µl containing 10 µM of each olionucleotide wera incubated for 10 min at 95°C and then cooled to room temperature. In the next step 5' overhanging ends were phosphorylated using T4 polynucleotide kinase (PNK, NEB). A 50 µl reaction contained 5 µl annealed oligonucleotide, 5 µl T4 DNA Kinase buffer (10x, NEB), 2 mM ATP and 2 µl T4 PNK. The reaction was incubated at 37°C for 30 min. 1 µl of the reaction was used for subsequent ligation. A 20 µl ligation reactions contained 25 ng linearized plasmid DNA, 1 µl of phosphorylated insert, 1 mM ATP, 2 µl T4 DNA Ligase buffer (10x, NEB) and 1 µl T4 DNA Ligase (NEB). Ligation was performed at 4°C over night. A ligation reaction without insert DNA served as a control. On the next day, E. coli Top10 cells were transformed with 10 µl of the ligation reaction. Sequences used are summarized in **Suppl. Tab5.2** and **Suppl. Tab5.5**.

4.2.4 Polymerase Chain Reaction (PCR)

For plasmids containing insert longer that 120 nt, inserts were amplified by PCR. For molecular cloning 100 µl reactions were performed using Q5 DNA polymerase (NEB). The reaction mix and PCR program is depicted in **Tab. 4.3**. For the purification of PCR reactions the Qiaquick PCR purification kit (Qiagen) was used, according to the manufacturers instructions. If necessary, PCR products were gel purified using the Qiaquick gel elution kit (Qiagen).

Table 4.3: Standard reaction mix and PCR program using Q5 DNA polymerase

Reation mix		**PCR program**		
1x	Q5 reaction buffer (NEB)	*Temperature*	*Time*	*Cycle*
30 pmol	forward oligonucleotide	98°C	3 min	
30 pmol	reverse oligonucleotide	98°C	30 s	
100 ng	template DNA	50°C	30 s	35x
20 nmol	dNTPs (Sigma-Aldrich)	72°C	15 s/kb	
1 µl	Q5 polymerase	72°C	7 min	
ad. 100 µl		8°C	∞	

Screening of colonies upon *E. coli* transformation for plasmids bearing DNA inserts of the correct size, colony PCR was performed. For pDLP-based plasmids the oligonucleotides luc2_3UTR_seq_fwd and luc2_seq_rev2 were used. For pCMV-based plasmids the oligonucleotides CMV_prom_fwd and SV40___polyA_rev. For pHDV-based plasmids the oligonucleotides pHDV_seq_fwd and AP1_pSP61 were used. Oligonucleotide sequences are listed in **Suppl. Tab. 5.7**. For colony PCR, a single colony was resuspended in 50 µl H_2O and 1 µl of that suspension used as template. The reaction mix and PCR program of a colony PCR is depicted in **Tab. 4.4**. In order to determine fragment size the reaction products were separated on 1 or 3% agarose gels and analyzed with a UV transilluminator.

4.2.5 Genomic Integration

For generation of stable cell lines, for the constitutive expression of mCherry and GFP with the *UCP3* wt repressive element and the double mutant MUTI/II, the Flp-In System (Thermo Fisher Scientific) was used. Therefore, HeLa HF1-3 cells were seeded to a density of 90000 cells in 12-well cell culture dish in growth medium without antibiotics. 24 h post seeding the cells were cotransfected with the pOG44 plasmid and the respective pFRT construct in a molar ratio of 9:1 using Lipofectamine 2000 (Thermo Fisher Scientific). A plasmid without FRT-site was cotransfected as negative control. 48 h post transfection growth medium was changed to selection medium supplemented with 15 µg/ml

Table 4.4: Reaction mix and PCR program used for colony PCR

Reaction mix		**PCR programm**		
1x	ThermoPol buffer	*Temperature*	*Time*	*Cycle*
7.5 pmol	forward oligonucleotide	96°C	4 min	
7.5 pmol	reverse oligonucleotide	96°C	30 s	
1 µl	cell suspension	50°C	30 s	25x
5 nmol	dNTPs	72°C	60 s/kb	
1.25 U	Taq polymerase	72°C	7 min	
ad. 25 µl		8°C	∞	

hygromycin B. During the selection process, the selection medium was exchanged every second or third day. After 2 weeks of selection no cells could be detected in the negative control. Successful integration of pFRT_UCP3wt and pFRT_UCP3_MUTI/II was analyzed by flow cytometry.

4.2.6 Flow Cytometry

For the analysis of the generated HeLa HF1-3 cells lines stably expressing GFP-UCP3 wt or MUTI/II fusions as well as the red fluorescent protein mCherry cells were seeded to a initial density of 200,000 cells in a 12-well cell culture dish. On the following day growth medium was discarded and the cells rinsed with PBS once. Cells were detached by adding 200 µl Trypsin-EDTA and incubation for 5 min at 37°C. After Trypsin treatment cells were diluted 1:1 with PBS and fluorescence at 510 nm and 610 nm was measured with a Beckman Coulter Cytoflex S. Analysis of flow cytometry data was performed with FlowJo (Version 10) software.

4.2.7 Luciferase Reporter Assay

To investigate the influence of different 3'UTR variants on luciferase activity a dual luciferase assay was performed. Dual luciferase assay of HEK293 cells was performed 24 h after transient transfection with pDLP-based reporter plasmids (see **4.2.1**) with the Dual-Glo® Luciferase Assay System (Promega) according to the manufacturer's instructions. First the cultivation medium was discarded and 100 µl growth medium without phenol red added to each well. For cell lysis, 100 µl of the Dual Glo® Luciferase reagent was pipetted to each well and incubated for 10 min at RT. After that the 200 µl cell lysate was transferred to a white 96-well Nunc™ MicroWell™ plate. Firefy luciferase activity was quantified by measuring the luminescence on an Infinite®M200 plate reader with an integration time of 1 s. Thereafter, 100 µl Dual Glo® Stop&Glo® reagent was added to each well. The reactions were incubated for 10 min before *Renilla* luciferase

activity was determined by measuring luminescence one more time. For the calculation of relative light units (RLU) the firefly luciferase values were first normalized to the *Renilla* luciferase values. After that mean and standard deviation were calculate from triplicates and normalized to the empty vector control.

4.2.8 RNA Extraction and Reverse Transcription

For total RNA extraction cells were first rinsed with 500 µl PBS. Then 400 µl Trizol was added to each well and the mixture was incubated for 5 min at RT. Cell lysis was accelerated by pipetting up and down several times. The Lysate was then transferred to a 1.5 ml reaction tube. Genomic DNA was separated by adding 80 µl chloroform. For efficient extraction, lysates were vortexed for 15 sec and then incubated for 3 min at RT prior to centrifugation for 15 min at 13,000 rpm and 4°C. The RNA containing aqueous phase was transferred into a fresh 1.5 ml reaction tube and 200 µl chloroform was added. After vortexing for 15 sec samples were again centrifuged for 15 min at 13,000 rpm and 4°C. Once again the aqueous phase was transferred into a fresh 1.5 ml reaction tube. For RNA precipitation, 200 µl ice cold isopropanol and 0.75 µl Glycoblue (Ambion) were added. After vigorous mixing samples were incubated for 10 min and centrifuged for 30 min at 13,000 rpm and 4°C. The supernatant was discarded and the pellet washed with 180 µl 70% EtOH. This was followed by a further centrifugation for 10 min at 13,000 rpm and 4°C. Then, the supernatant was discarded and the pellet was air dried for 5 min at RT. The pellet was then dissolved in 35 µl H_2O. In order to eliminate DNA contamination from RNA preparations, a DNase digestion was performed. Therefore, 4 µl 10× TurboDNase buffer and 1 µl TurboDNase were added to RNA samples and incubated for 15 min at 37°C. The reaction was incubated for 15 min at RT. For RNA precipitation, 4 µl 3 M sodium acetate (pH=6.5) was added and incubated for 1 h at -20°C. The RNA pellet was rinsed with 70% (v/v) ethanol once and then air dried for 5 min at RT. After isolation, 500 ng of RNA was quality checked on a 1% agarose gel.

For mRNA analysis, 1 µg total RNA was reverse transcribed to cDNA using MuLV reverse transcriptase (Thermo Fisher Scientific) using random hexamer oligonucleotides. Therefore, 1 µg total RNA was diluted in 10 µl H_2O and mixed with 30 µl reaction mixture. The reaction mixture and the PCR program for the reverse transcription are depicted in **Tab. 4.5**.

After reverse transcription, 160 µl H_2O was added to the reaction, resulting in a final cDNA concentration of 5 ng/µl.

For the detection of miRs reverse transcription with miR specific stem-loop oligonucleotides was employed. Conventional PCR and qPCR methods require DNA matrices with a length of ≥40 nt so that the specific forward and reverse oligonucleotides can hybridize with the DNA matrices. In stem-loop PCR, reverse transcription is performed with a special oligonucleotide (given in **Suppl. Tab. 5.7**) that artificially extends the cDNA. The oligonucleotide consists of a stable stem-loop structure and the last six complementary nucleotides from the 3' end of the miR to be detected [140]. For

Table 4.5: Reaction mix and PCR program for reverse transcription with MuLV and random hexamers

Reaction mix		**PCR program**	
8 µl	$MgCl_2$ (25 mM)	*Temperature*	*Time*
4 µl	10 x transcription buffer	20°C	10 min
0,4 µl	dNTPs (100 mM)	42°C	15 min
2 µl	random hexamers (0.2 ng/µl)	99°C	5 min
2 µl	RNase inhibitor Ribolock	4°C	∞
2 µl	MuLV reverse Transcriptase		
ad. 30 µl			

miR detection, 100 ng total RNA was transcribed into cDNA using Supersript II reverse transcriptase. For normalization in subsequent qPCR experiments, the small nucleolar RNA U48 was co-transcribed into cDNA. For the reaction, 4 µl total RNA was mixed with 6 µl of the reaction mix. The reaction mix and PCR program for stem-loop reverse transcription is depicted in **Tab. 4.6**.

Table 4.6: Reaction mix and PCR program for reverse transcription with Superscript II and stem-loop oligonucleotides

Reaction mix		**PCR program**		
1 µl	stem-loop oligonucleotide (mir-X), 500 nM			
1 µl	stem-loop oligonucleotide (U48), 500 nM	*Temperature*	*Time*	*Cycle*
1 µl	DTT, 100 mM	60°C	30 min	
0.25 µl	dNTPs, 10 µM	30°C	30 s	
0,25 µl	Superscript II reverse transcriptase	42°C	30 s	60x
		50°C	60 s	
2 µl	5x First Strand PCR buffer	70°C	15 s	
0.5 µl	H_2O	8°C	∞	
6 µl + 4 µl RNA (25 ng/µl)				

After the stem-loop PCR, 40 µl H_2O was added to the reaction resulting in a final cDNA concentration of 2 ng/µl.

4.2.9 Quantitative RT-PCR

All quantitative real-time PCR (qPCR) analysis were performed on a StepOnePlus™ Real-Time PCR System (Thermo Fischer Scientific). SYBR green intercalates into the major grove of double-stranded DNA and fluoresces in the bound state. The increase in amplification of the target cDNA was measured by the increase of fluorescence and therefore allows to quantify the relative amount of a target RNA compared to a control. For quantification of RNAs in HEK293, HeLa HF1-3 cells and HUVEC *RPLP0* mRNA was used for normalization. For C2C12 RNA *GAPDH* mRNA was used for normalization. All oligonucleotides used in qPCR experiments are listed in **Suppl. Tab.5.7**. First a master mix containing Fast SYBR green Master Mix (Thermo Fischer Scientific) and a target specific oligonucleotide pair was prepared. 15 µl of the master mix were pipetted into a 96-well qPCR plate. After that cDNA was added. The plate was sealed with an optical cover centrifuged shortly to mix cDNA and reaction mix. The reaction mix and qPCR program used are depicted in **Tab. 4.7**. Each oligonucleotide pair was tested to produce only the desired product by sequencing of the PCR product.

Table 4.7: qPCR reaction mix and PCR program for SYBR green based qPCR

Reaction mix		**PCR program**		
		Temperature	*Time*	*Cycle*
10 µl	SYBR Green Master Mix	95°C	20 s	
5 µl	oligonucleotide mix (10 µM each) containg forward and reverse oligonucletides	95°C	3 s	
		60°C	30 s	40x
5 µl	cDNA (5 ng/µl)	95°C	15 s	
		60°C	1 min	
		heat in 0.3°C steps	to 95 °C	
		95°C	15 s	

For the detection of miRs, TaqMan qPCR was performed. This method is based on a probe, which carries a fluorescent dye and a quencher, that can bind specifically to the target cDNA. Upon amplification the polymerase cleaves the probe and releases the dye which results in a specific emission. The probe binding site was introduced via the stem loop oligonucleotide in cDNA synthesis (see 4.6).The reaction mix and qPCR program are depicted in **Tab. 4.8**. Oligonucleotides are given in **Suppl. Tab.5.7**.

4.2.10 mRNA Decay Assay

For mRNA decay experiments, 200,000 HF1-3 cells stably expressing either *UCP3* wt or *UCP3* MUTI/II, were seeded in 12-well cell culture dishes. 24 h after seeding the cells, the medium was changed to DMEM supplemented with 5 µg/ml actinomycine D

Table 4.8: Reaction mix and PCR program for TaqMan based qPCR of miRs

Reaction mix		**PCR program**		
1 µl	universal oligonucleotide (10 µM)			
1 µl	forward oligonucleotide (10 µM)	*Temperature*	*Time*	*Cycle*
1 µl	universal probe 21	50°C	2 min	
10 µl	2x TaqMan Fast Universal Master-mix	95°C	10 min	
		95°C	15 s	
2 µl	H_2O	60°C	1 min	40x
15 µl + 5 µl RNA (25 ng/µl)				

(Sigma) for 30 min. After that medium as changed back to standard growth medium. Cells were harvested after 0, 2, 4 and 6 h and total RNA was extracted using Trizol. After reverse transcription, *GFP* mRNA levels were determined by SYBR green based qPCR analysis (see **4.2.9** . For the determination of mRNA half-life, *GFP* mRNA levels were normalized to *RPL0* mRNA levels and plotted against time. mRNA half-lives were calculated assuming a first order decay rate. MRNA half-live was calculated by the equation:

$$Y = (Y_0 - Plateau) * e^{(-K+X)+Plateau} \tag{4.1}$$

4.2.11 Western Blot Analysis

For the preparation of whole cell protein lysates, cells were lysed in lysis buffer [137 mM NaCl, 10% glycerol, 20 mM Tris-HCl pH 8.0, 2 mM EDTA pH 8.0, 1% Igepal, 5 µl protease inhibitor cocktail (Sigma) for 20 min on ice. In order to separate soluble and insoluble fraction, cell lysates were centrifuged for 15 min with 17,000 g at 4°C. The supernatant was transferred to a fresh 1.5 mM reaction tube and protein content was determined according to the Bradford method. For SDS-PAGE, the Mini-PROTEAN® TGX Stain-Free™ anyKd gels (Bio-Rad) were used. 10 µg of protein sample were mixed with protein loading buffer (PAP) (12.5% glycerol, 1% SDS, 3.75% 2-Mercaptoethanol, Bromophenol blue) and incubated for 10 min at 95°C. SDS-PAGE was run at 180 V in 1x Laemmli buffer. Prior to blotting, total lane protein was visualized as loading control with the ChemiDoc Imaging System (Bio-Rad). Proteins were transferred onto PVDF membranes (Bio-Rad) using the Trans-Blot® Turbo™ transfer system (Bio-Rad). Westernblot membranes were blocked with TBST containing 5% BSA (Applichem) overnight at 4°C. On the next day westernblot membranes were incubated with the primary antibody Roquin-1/-2 (Millipore 3F12) with agitation for 1 h at RT. Anti-Roquin was diluted 1:1000 in TBST. Afterward membranes were washed 3x with TBST with agitation for 5 min. The secondary antibody Horseradish peroxidase-conjugated anti-rat IgG

(Jackson Immunoresearch) was applied at a dillution of 1:10,000 in TBST for 1 h with agitation. Blots were washed 3x as described above. Blots were developed with the ECL select (Life technologies) reagent. Imaging was performed with the ChemiDoc Imaging System.

4.2.12 *In Vitro* Transcription

For RNA affinity purification the *UCP3* wt, MUTI, MUTII, MUTI/II and R1 RNAs were synthesized by *in vitro* transcription. Likewise CDEImut, CDEIImut and CDEI/IImut used in EMSA experiments were synthesized by *in vitro* transcription. To generate linear DNA templates the corresponding pHDV constructs were digested with HindIII which cleaves the pHDV plasmid directly 3' to the HDV ribozyme. Linearized plasmid DNA was purified by phenol extraction prior to *in vitro* transcription. Transcription was performed overnight at 37°C. The reaction mix is depicted in **Tab. 4.9**. After transcription 1/2 V 0.5 M EDTA pH 8.0 was added to the reaction. Then the RNA was precipitated with ethanol. After precipitation, the RNA pellet was dissolved in formamide loading buffer (Formamide, 0.25 mM EDTA, Bromphenol Blue) and the reaction products were separated on a 6% PAA gel. Bands were visualized by UV-shadowing. The band of interest was was excised and the gel pieces transferred into a 50 ml reaction tube. The RNA was eluted with 10 ml 300 mM NaAc pH 6.5 at 4°C overnight wit continuous agitation. On the next day, gel debris were separated by filtration, the RNA ethanol precipitated and the pellet dissolved in an appropriate amount of H_2O. The RNA concentration was determined by Nanodrop.

Table 4.9: *In vitro* transcription reaction mix using T7 RNA ploymerase

Amount	Component
200 µg	DNA template
80 µl	25 mM $Mg(Ac)_2$
200 µl	1 M Tris-HCl pH 8.0
20 µl	1 M DTT
10 µl	200 mM Spermidine
40 µl each	100 mM ATP, UTP, GTP, CTP
5 µl	T7 RNA polymerase
ad. 1 ml	H_2O

4.2.13 RNA Affinity Purification

For purification of *UCP3* binding proteins, the purified RNA was biotinylated at the 5'-end in a three step reaction. First 600 pmol RNA were dephosphorylated using calf

intestine phosphatase (CIP, Roche). The reaction mix is depicted in **Tab. 4.10**. The reaction was incubated for 30 min at 37°C.

Table 4.10: Reaction mix for the dephosphorylation of RNA using CIP

Amount	Component
600 pmol	RNA
1 µl	10x CIP buffer
1 µl	CIP
ad. 10 µl	H_2O

In the second step γ-S-ATP (Biomol) was used to transfer an sulfhydroxyl group to the dephosphorylated 5' end of the RNA using T4 polynucleotid kinase (T4 PNK, NEB). Necessary reaction compounds were added directly to the CIP reaction. The phosphorylation was incubated for 30 min at 37°C. The reaction mix is depicted in **Tab. 4.11**.

Table 4.11: Reaction mix for γ-S-ATP phosphorylation of RNA using PNK

Amount	Component
10 µl	CIP RNA
0.2 mM	γ-S-ATP
2 µl	10x T4 PNK buffer
2 µl	T4 PNK
ad. 20 µl	H_2O

In the last step 10 µl Biotin-long-arm maleimide (Vector Laboratories) was added and incubated for 30 min at 65°C. Unincorporated label was depleted by LiCl precipitation. Therefore, 70 µl H_2O was added to the reaction mixture. Then, 12.5 µl 4 M LiCl and 375 µl ethanol were added. After vigorous mixing samples were incubated for 1 h at -80°C. Subsequently the precipitation mixture was centrifuged for 10 min at 4°C. The supernatant was discarded and the RNA pellet was rinsed with 500 µl 70% (v/v) ethanol once. After subsequent centrifugation the pellet was air dried and dissolved in 20 µl H_2O. The RNA concentration was determined by Nanodrop.

For affinity purification 100 µl Dynabeads M-280 (Invitrogen) were washed 5x with 1 ml incubation buffer (10 mM Tris-HCl pH 7.4, 150 mM KCl, 0.5 mM DTT, 0.05% Igepal, 100 U/ml Ribolock). HEK293 cell lysates were prepared as follows: Cells were rinsed once with PBS before Trypsin-EDTA digestion at 37°C for 3 min. The Reaction was stopped by the addition of DMEM medium. Then, cells were transferred to a 50 and centrifuged. The supernatant was discarded and the cells were resuspended in 150 µl (per T75 culture flask) lysis buffer (10 mM Tris-HCl pH 7.4, 150 mM KCl, 0.5 mM DTT,

0.5% Igepal, 100 U/ml Ribolock). Subsequently, the cells were snap freezed in liquide nitrogen and then incubated on ice for 20 min. Then, the lysates were centrifuged for 10 min at 13,000 rpm. The supernatant was transferred to a fresh reaction tube and the protein concentration was determined according to the Bradford method. For biotin-avidin based RNA affinity purification 200 pmol RNA was then conjugated to Dynabeads M-280 (Invitrogen) in 500 µl incubation buffer for 2 h at 4°C with continuous rotation. Then 1 mg whole cell protein of HEK293 cells together with 200 µg yeast tRNA (Sigma Aldrich) and 5 mg Heparin (Sigma Aldrich) was added to the beads and incubated for 1 h at 4°C followed by 15 min at room temperature with continuous rotation. Afterward beads were washed 5x with 1 ml incubation buffer, resuspended in 30 µl PAP (see section 4.2.11), and boiled at 95°C for 10 min. Eluted proteins were separated by SDS-PAGE (BioRad).

4.2.14 Mass Spectrometry

Proteins from RNA affinity purification were analyzed by mass spectrometry as previously described in [37]. Analysis was performed by Hennig Urlaub and Uwe Plessman (MPI für Biophysik, Göttingen). In contrast to the procedure described in [37] the protein samples were separated on a NuPAGE SDS-PAGE gel. After gel electrophoresis the entire lanes were cut into 23 slices and further processed as described. The LC-MS analyses were performed on a Q-Exactive Plus MS (Thermo Fisher Scientific) coupled to an UltiMate 3000 HPLC (Thermo Fisher Scientific) under similar condition as described except that LC separation time was 48 min. The MS data was acquired by scanning the precursors in mass range from 350 to 1600 m/z at a resolution of 70,000 at m/z 200. The top 20 precursor ions were chosen for MS2 by using data dependent acquisition (DDA) mode at a resolution of 15,000 at m/z 200 with maximum IT of 50 ms. The generated data was then analyzed as previously described using Mascot (Version 2.3.02) as search engine. Ions were searched against the against Uniprot database (Taxonomy human, 155190 entries, 20161124). The data was than evaluated with the Scaffold software (Version 4.8.2). For further analysis a cut-off of at least five unique peptides detected in at least one of the samples was used.

4.2.15 In-Line Probing Analysis

In-line probing was performed as previously described [205]. First, *UCP3* wt RNA was synthesized by *in vitro* transcription using T7 RNA polymerase (see section **4.2.12**). In the next step, *UCP3* wt RNA was dephosphorylated using CIP (Roche). The reaction mix is depicted in **Tab. 4.12**. The dephosphorylation reaction was incubated at 55°C for 30 min. After that, the reaction was incubated at 95°C for 10 min to inactivate the enzyme.

In the next step the RNA was 5' radioactively labeled using PNK (Roche). The reaction mix is depicted in **Tab. 4.13**. The phosphorylation reaction was incubated at

Table 4.12: Reaction mix for the 5' dephosphorylation using CIP

Amount	Component
20 pmol	RNA
15 µl	H_2O
2 µl	10 x CIP buffer (Roche)
1 µl	CIP

37°C for 1 h. After that 20 µl 2x loading buffer (10 M urea) was added to the sample and the sample subjected to denaturing 6% PAA gel electrophoresis. Bands were visualized by exposing the gel to X-ray film. Then, the RNA was excised and the gel slice transferred to a fresh 1.5 ml reaction tube. The 5'[γ-^{32}P]-labeled RNA was recovered by gel extraction with chrush/soak solution (200 mM NaCl, 10 mM Tris–HCl pH 7.5, 1 mM EDTA) and subsequent ethanol precipitation. Radioactivity was quantified with a liquid scintillation counter.

Table 4.13: Reaction mix for the [γ-^{32}P]-ATP phosphorylation using PNK

Amount	Component
10 µl	dephosphorylated RNA
1 µl	H_2O
2 µl	10 x PNK Puffer
6 µl	[γ-^{32}P]-ATP
1 µl	PNK

For in-line probing analysis of the 64 nt *UCP3* wt RNA, 40 kcpm of the [γ-^{32}P]-labeled RNA were incubated at 22°C for 40 h in in-line reaction buffer (10 mM Tris-HCl pH 8.3, 10 mM $MgCl_2$, 100 mM KCl) in a 20 µl reaction. To generate a size marker, the [γ-^{32}P]-labeled RNA was subjected to alkaline hydrolysis. Therefore the RNA was incubated in 50 mM Na_2CO_3, pH 9.0 at 96°C for 3 min. To map guanines, the [γ-^{32}P]-labeled RNA was incubated for 3 min at 55°C with 20 U RNase T1 (Ambion). The reaction mixes for the alkaline hydrolysis and RNase T1 treatment are depicted in **Tab. 4.14**. After in-line reaction, alkaline hydrolysis or RNase T1 treatment, samples were precipitated with 60 µl acetone:ethanol (1:1), 2 µl 3 M sodium acetate and 1 µl Glycoblue (Life Technologies) in liquid nitrogen. Pellets were rinsed with 70% ethanol once and dissolved in 55 M urea. The reaction products were then separated by denaturing 10% PAGE. After electrophoresis, gels were dried on a vacuum slab gel drier at 80°C for at least 1 h. Dried gels were exposed to a phosphor imaging screen over night and analyzed using Phosphoimager FL-5000 (Fuji).

Table 4.14: Reaction mix for alkaline hydrolysis and RNase T1 treatment

alkaline hydrolysis		**RNase T1 treatment**	
17 µl	H_2O	2 µl	Na-Citrat 250 mM (pH 9.0)
40 kcpm	[γ-^{32}P]-RNA	40 kcpm	[γ-^{32}P]-RNA
2 µl	Na_2CO_3 (500 mM,pH 9.0, 10 mM EDTA)	2 µl	RNase T1 (10 U/µl)
add. 20 µl	H_2O	add. 15 µl	8 M urea

4.2.16 Electrophoretic Mobility Shift Assay (EMSA)

Mouse Roquin-1 core ROQ domain and N-terminus protein fragments as well as the corresponding K220A, K239A, R260A (AAA) mutants [59] were provided Dr. Adreas Schlundt (Johann Wolfgang Goethe-Universität, Frankfurt). EMSAs were performed as described in ([59]. *UCP3* wt, CDEImut, CDEIImut and CDEI/IImut RNAs were prepared by 5' end labeling of 10 pmol *in vitro* transcribed RNA with [γ-^{32}P]-ATP using CIP (Roche) as described in section **4.2.15**. Labeled RNA was purified from unincorporated nucleotides by ammonium acetate (Roth) precipitation. Therefore, 10 µl 7.5 M ammonium acetate and 80 µl ethanol were added to the reaction mix. After vortexing the reaction was incubated on ice for 10 min and then cetrifuged at RT and 17,000 g for 10 min. The supernatant was discarded and the pellet rinsed twice with 70% (v/v) ethanol. The RNA pellet was air dried for 5 min and resolved in 50 µl H_2O. Radioactivity was quantified with a liquid scintillation counter. Prior to binding reactions a mastermix containing RNA (40 kcpm per 20 µl reaction), 2.5 µg/µl yeast tRNA (Sigma Aldrich), 20 mM HEPES pH 7.4, 50 mM NaCl, 1 mM $MgCl_2$, 1 mM DTT and 1 µg/µl BSA was prepared. A dilution series of 10x protein stocks was prepared with the same buffer components. For each binding reaction, 5 µl of the 10x protein stock was added to 15 µl of the mastermix. The binding reactions were incubated for 20 min at RT. After addition of 4 µl 30% (v/v) glycerol as loading buffer, the RNP complexes were resolved by non-denaturating PAGE (6% polyacrylamide, 5% glycerol, and 0.5x TBE) at 120 V for 40 min at RT. Subsequently gels were dried at 80°C for 30 min on a vacuum gel drier and then exposed to a phosphor imager screen. For data analysis the fraction of Roquin-bound RNA was plotted against the corresponding Roquin concentration. Using Graphpad Prism software, a curve with the following equation

$$y = B_{max} \times \frac{X^h}{K_D^h + X^h} \tag{4.2}$$

was fitted to the data points, where B_{max} is the maximum specific binding, K_D the dissociation constant, and h the Hill coefficient.

4.2.17 siRNA-mediated Depletion of Roquin Proteins

100,000 HEK293 or 200,000 HF1-3 cells were reverse transfected in a 12-well cell culture dish with either siCTRL or siROQ1/2 mix at a final concentration of 100 nM using Lipofectamine RNAiMAX (Thermo Fischer Scientific) and OptiMEM (Sigma Aldrich) according to the manufacturer's instructions. RNA and protein samples were prepared 48 h post transfection. To investigate the influence of Roquin knockdown on luciferase activity, a transient transfection of the reporter plasmid was performed 24 h after reverse transfection of siRNAs. For knockdown of Roquin proteins in HUVEC 350,000 HUVECs were seeded in 60 mm cell culture dishes. 24 h after seeding, cells were forward transfected with either siCTRL or siROQ1/2 mix at a final concentration of 100 nM using Lipofectamine RNAiMAX (Thermo Fisher Scientific) according to the manufacturers instructions. RNA and protein samples were prepared 48 h post transfection. For knockdown of Roquin proteins in C2C12 cells, 50,000 cells were reverse transfected in a 12-well cell culture dish with 100 nM siCTRL or mmsiROQ1/2 mix. 48 h post transfection cells had reached 90-100% confluency. To induce differentiation the medium was switched to a culture medium containing 2% horse serum. After 24 h of differentiation RNA and protein samples were prepared. All siRNAs used in this study were synthesized by Sigma-Aldrich and correspond to the sequences (sense strand) given in **Tab. 4.15**.

Table 4.15: siRNAs used in this study

Name	Target	Sequence
siCRTL	control	5'-UUCUCCGAACGUGUCACGU[dT][dT]-3'
siROQ1	human Roquin-1	5'-CCAAGAAAUGUGUAGAAGA[dT][dT]-3'
siROQ2	human Roquin-2	5'-UCUUCUACACAUUUCUUGG[dT][dT]-3'
mmsiROQ1	mouse Roquin-1	5'-CGCACAGTTACAGAGCTCA[dT][dT]-3'
mmsiROQ2	mouse Roquin-2	5'-GGACTTGGCTCATAAATCA[dT][dT]-3'

4.2.18 Computational Methods

All bioinformatic analyses were carried out by Dave Mathews (Rochester University, New York, USA) and coworkers. The methods are as described in [113].

Search for conserved structures using Dynalign

For the bioinformatic prediction of structurally conserved RNA elements human UTR sequences (hg18) were downloaded from the UCSC genome browser (`http://hgdownload.soe.ucsc.edu/downloads.html#human`). Next the acquired UTRs were fragmented into overlapping windows of 100 nt with a step size of 50 nt. Then, the corresponding mouse

sequences were identified using the human-mouse pairwise genomic alignment obtained from UCSC genome browser (`http://hgdownload.soe.ucsc.edu/goldenPath/hg18/vsMm9/`). Each pair of 100 nt long human and mouse sequences was input to the *Dynalign* algorithm for RNA structure prediction [206]. To evaluate the results by their probability of being a conserved structure Dynalign/SVMz was used [135].

"First, a measurement of the significance of the calculated free energy changes, z-score, was calculated as compared to shuffled sequences [122]. The z-score, the number of standard deviations from the mean, is calculated as:

$$z = \frac{m - \mu}{\sigma} \tag{4.3}$$

where m is the predicted minimum free energy change of the input, μ and σ is the mean and the standard deviation of free energy changes of the shuffled sequences. Second, the Structure Conservation Index (SCI) was calculated (4):

$$SCI = \frac{\Delta G^0_{dynalign}}{\Delta G^0_{seq1} - \Delta G^0_{seq2}} \tag{4.4}$$

where $\Delta G^\circ_{\text{dynalign}}$ is the Dynalign predicted total minimum free energy change of the two sequences and $\Delta G^\circ_{\text{seq1}}$ and $\Delta G^\circ_{\text{seq2}}$ are the predicted minimum free energy change for each sequence alone [116]. The SCI is an indicator of structural conservation. The larger the SCI, the more conserved the secondary structures are. Using the stability z-score and the SCI, with the nucleotide composition of the two input sequences a support vector machine, called SVMz, was trained to classify the probability that a window is a conserved structure. The windows of sequence pairs were ranked by their probability of being a conserved structure and the windows with a probability greater than 0.9 were further analyzed." [113]

Prediction of human CDEs and their folding probability

"The whole human genome sequence (hg19) was downloaded from UCSC and the sequence was processed to 3'UTRs using bam2x (`https://github.com/nimezhu/bam2x`). RNAmotif, which takes an RNA descriptor, an RNA sequence file and outputs the positions where the RNA motif is found [207], was run to find genes that contain the desired motif. The motif is an RNA stem-loop, which has a stem length of 6-8 bp and a triloop with sequence 5'-YRN-3'. Then, rm2ct from the RNAmotif program was used to convert an output file to a ct file, which indicates the sequence and base pairing. The efn2 program from RNAstructure [208], which takes a ct file and calculates the folding free energy change , was run and the equilibrium constant for the interior part of the desired motif is calculated using the following equation:

$$K_{\text{interior}} = e^{-\frac{\Delta G^0_{37}}{RT}} \tag{4.5}$$

here, R is the gas constant and T is the absolute temperature in K.

The probability of a certain motif occurring in a given RNA sequence is as follows:

$$P_{\text{motif}} = K_{\text{interior}} * Q_{\text{exterior}}/Q \tag{4.6}$$

here, Q_exterior is the partition function over configurations outside the motif and Q is the total partition function [151].

In order to calculate the exterior probability, a C++ program called ProbStemloop, was written using RNAstructure. ProbStemloop takes three input parameters: the sequence file, the starting position of a stem-loop, the length of the stem-loop and outputs the exterior probability for the motif. The sequences processed by the ProbStemloop were truncated to 400 nt up- and downstream. Prior work demonstrated this much sequence is a sufficient representation of the local folding in mRNA [150]. Finally, the probability of motif is calculated with equation **4.6**. All the code, written in python and C++, is available at `https://github.com/hongyingsun1101/Motif-Probability`. Similarly, stem-loops containing a hexaloop were searched across the human genome and the probability of this hexaloop containing stem-loops was also calculated. The motif is an RNA stem-loop which has a stem with a length of 6-8 bp and a hexaloop with the sequence 5'-GUUYUA-3'." [113]

Evolutionary conservation of putative CDEs and ADEs

"The RefSeq genomic coordinates and sequences for mRNA 3'UTRs containing predicted CDEs were obtained from UCSC Table browser [209] using the hg19/GRCh37 reference genome. Because of occasional inconsistencies in UCSC and NCBI RefSeq genomic coordinates for 3'UTRs, UCSC nucleotide sequences were scanned to identify the respective CDE and genomic coordinates of the CDEs were adjusted as needed. Multiz 46-way [210] multiple species alignments across each chromosome were obtained from UCSC golden path database (updated 30-Oct-2009). For each instance of the CDE motif, we determined conservation using multiple sequence alignments that fully span the length of the motif. We checked for exact sequence conservation of each hg19 CDE across *Mus musculus* (mm9, July 2007), *Pan troglodytes* (panTro2, Mar 2006), *Bos taurus* (bosTau4, Aug 2006), and *Canis familiaris* (canFam2, May 2005). In the cases where there are gaps at the 5' or 3' end of a CDE in any of the species, the nucleotides 5' or 3' to the gap, respectively, were used to fill in the gap. In addition, we checked for conservation across the five species allowing tolerated changes. Tolerated changes were defined as any of the following: (a) changes in the triloop that still fit the definition of the motif (5'-stem-YRN-stem-3'), (b) changes to the length of the stem such that only 6 base pairs are required to be conserved, and changes to the stem that can still form canonical base pairs. These could be (c) compensating base pair changes, or (d) G ∘ U wobble base pairs."[113]

Abbreviations

The following abbreviations were used in this thesis:

Table 4.16

Abbreviation	Full word
A	adenine
actD	actinomycin D
ADP	Adenosindiphosphate
ARE	AU-rich elemet
ARE-BP	AU-rich element binding protein
AS	aternative splicing
ATP	adenosintriphosphate
AUF1	AU-rich binding factor
BP	branch point
bp	base pair(s)
C	Cytosine
CBC	cap binding complex
CDE	constitutive decay element
CDS	coding sequence
cDNA	complementary/copy DNA
cpm	counts per minute
CPSF6	cleavge and plolyadenylation specific factor 6
DANN	desoxy ribonucleic acid
DHX	DEAD bos helicase domain
DHX	DED box helicase
DMEM	Dulbecco modified eagle medium
dNTP	desoxy-nucleotidtriphosphate
dsRNA	double stranded RNA
EIF4E	translation initiation factor 4E
Exp5	Exportin
FMRP	fragile X mental retardation protein 1
G	Guanine
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GFP	*green fluorescent protein*
hnRNP	heterogenous ribonucleoprotein

Abbreviation	Full word
HuR	human antigen R
ICOS	inducible T cell co-stimulator
IDD	intrinsically disordered domain
IRE	Iron-responsive elemet
IRES	internal ribosomal entry site
IRP	IRE binding protein
kDa	kilo Dalton
KH	K-homology domain
KSRP	K-homology splicing regulatory protein
miR	micro RNA
mRNA	messenger RNA
mRNP	messenger ribonucleoprotein particle
nc	non coding
ncRNA	non coding RNA
NGD	no-go decay
NMD	nonsense-mediated decay
NMR	nuclear megnetic resonance
NSD	non-stop decay
nt	nucleotide(s)
OH	Hydroxyl
PABP	poly(A) binding protein
PAGE	Polyacrylamide gel electrophoresis
PAN	poly(A)-specific ribonuclease subunit
PARN	poly(A)-specific ribonuclease
P-body	processing body
PCR	polymerase chain reaction
piRNA	Piwi -imteracting RNA
pri	primary
PTC	pre-mature stop codon
PTR	post transcriptional regulation
qPCR	quantitative PCR
R	Purine
RBD	RNA binding domain
RBP	RNA binding protein
RING	really interesting new gene
RISC	RNA induced silencing complex
RLU	relative light units
RNA	ribonucleic acid
RNase	ribonuclease
RPLP0	ribosomal protein lateral stalk subunit P0
RRM	RBA recognition motif

Abbreviation	Full word
RT	reverse transcription
RT-qPCR	reverse transcriptase qPCR
SECIS	seleno cycsteine insertion element
SELEX	systematic evolution of ligands by exponential enrichment
siRNA	small interfering RNA
SMD	Staufen-mediated decay
snRNP	small nucleolar ribonucleoprotein
SR	serin/arginie
SS	splice site
TIA-1	T cell intracellular antigen-1
TIAR	TIA-realted
TNF	tumor necrosis factor a
TTP	tristetraprolin
U	Uracil
UCP3	uncoupling protein 3
uORF	upstream open reading frame
UTR	untranslated region
UV	ultra violet
wt	wilde type
Xrn1	5'-3' Exoribonuclease 1
Y	Pyrimidine
ZnF	zinc finger domain

Publications

Data from this thesis was published in:

1. **Johannes Braun**, Sandra Fischer, Zhenjiang Z Xu, Hongying Sun, Dalia H Ghoneim, Anna T Gimbel, Uwe Plessmann, Henning Urlaub, David H Mathews, Julia E Weigand; *Identification of new high affinity targets for Roquin based on structural conservation*, Nucleic Acids Research **2018**.

Other publications:

2. Jeannine Jaeger, Florian Groher, Jacqueline Stamm, Dieter Spiehl, **Johannes Braun**, Edgar Dörsam and Beatrix Suess; *Characterization and inkjet printing of RNA-aptamer to create a paper-based biosensor for ciprofloxacin based on fluorescence changes*, submitted.

3. Boussebayle Adrien, Torka Daniel, Ollivaud Sandra, **Johannes Braun**, Groher Florian and Beatrix Suess; *Development of a paromomycin synthetic riboswitch using a RNA based capture-SELEX*, submitted.

4. Florian Groher, Cristina Bofill-Bosch, Christopher Schneider, **Johannes Braun**, Sven Jager, Katharina Geißler, Kay Hamacher, Beatrix Suess; *Riboswitching with ciprofloxacin - development and characterization of a novel RNA regulator*, Nucleic Acids Research 46(4) 2121–2132, (**2018**).

5. Patrick Rühl, Uwe Pöll, **Johannes Braun**, Andreas Klingl, Arnulf Kletzin; *A Sulfur Oxygenase from the Haloalkaliphilic Bacterium Thioalkalivibrio paradoxus with Atypically Low Reductase Activity*, Journal of Bacteriology 199(4) e00675-16, (**2017**) .

6. Michael M. Kunze, Fabienne Benz, Thilo F. Brauß, Sebastian Lampe, Julia E. Weigand, **Johannes Braun**, Florian M. Richter, Ilka Wittig, Bernhard Brüne, Tobias Schmid; *sST2 translation is regulated by FGF2 via an hnRNP A1-mediated IRES-dependent mechanism*, Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms 1859(7) 848-859, (**2016**).

Acknowledgements

I would like to thank

Dr. Julia Weigand, für die Bereitstellung des spannenden Projekts sowie die hervorragende Betreuung, Deine Unterstützung und Geduld sowie das große Vertrauen das Du mir in meinem Arbeiten immer gewährt hast. Dein wissenschaftlicher Enthusiasmus und Dein Glaube in meine Arbeit haben mich immer wieder motiviert.

Prof. Beatrix Süß, für die Möglichkeit meine Arbeit in Ihrem tollen Labor durchzuführen, und an der erfolgreichen Verwirklichung vieler Projekte beteiligt zu sein.

Prof. Viktor Stein, danke ich recht herzlich für die freundliche Übernahme des Zweitgutachtens.

Dem ganzen Arbeitskreis Süß, für die gute Stimmung im Labor, die Kollegialität und Hilfsbereitschaft und die lustige und abwechslungsreiche Zeit.

Prof. David M. Mathwes, **Honying Sun**, **Dahlia Gohneim**, for the fruitful collaboration and all the computing. The input you provided was indispensable and really pushed the project.

Prof. Henning Urlaub, für die gute Zusammenarbeit und die LC-MS/MS Analyse ohne die wir nicht weitergekommen wären.

Dr. Florian Groher, für die vielen wissenschaftlichen Diskussionen, kritischen Fragen und Freitag-Meetings.

Sandra Fischer, für vielen Stunden HUVEC Zellkulultur und **Theresa Gimbel** für ihre Unterstützung.

Britta Kluge, für die experimentelle Unterstützung und ihre positive Art.

Britta Schreiber, für Mittagspausen außerhalb des Hühnerstalls und die Kaffepausen, die immer eine willkommene Abwechslung waren.

meiner **Familie**, die mich immer unterstützt hat und mir dadurch überhaupt erst ermöglicht hat diesen Weg zu bestreiten.

meiner Frau **Yasmin** und meinen beiden Mädels **Emmi** und **Lina**. Danke, für all die Entbehrungen die ihr auf euch genommen habt und das ihr mich trotzdem zu jeder Zeit bedingungslos unterstützt habt. Ohne euren Rückhalt wäre ich nicht, wo ich heute stehe.

Bibliography

[1] Gehring, N. H., Wahle, E. & Fischer, U. Deciphering the mRNP Code: RNA-Bound Determinants of Post-Transcriptional Gene Regulation. *Trends in Biochemical Sciences* **42**, 369–382 (2017).

[2] Iyer, B. V., Kenward, M. & Arya, G. Hierarchies in eukaryotic genome organization: Insights from polymer theory and simulations. *BMC Biophysics* **4** (2011). `https://doi.org/10.1186/2046-1682-4-8`.

[3] Venkatesh, S. & Workman, J. L. Histone exchange, chromatin structure and the regulation of transcription. *Nature Reviews Molecular Cell Biology* **16**, 178–189 (2015).

[4] Van Steensel, B. Chromatin: Constructing the big picture. *EMBO Journal* **30**, 1885–1895 (2011).

[5] Maston, G. A., Evans, S. K. & Green, M. R. Transcriptional Regulatory Elements in the Human Genome. *Annual Review of Genomics and Human Genetics* **7**, 29–59 (2006).

[6] Halbeisen, R. E., Galgano, A., Scherrer, T. & Gerber, A. P. Post-transcriptional gene regulation: From genome-wide studies to principles. *Cellular and Molecular Life Sciences* **65**, 798–813 (2008).

[7] Moore, M. J. From birth to death: The complex lives of eukaryotic mRNAs. *Science* **309**, 1514–1518 (2005).

[8] Müller-Mcnicoll, M. & Neugebauer, K. M. How cells get the message: Dynamic assembly and function of mRNA-protein complexes. *Nature Reviews Genetics* **14**, 275–287 (2013).

[9] Ramanathan, A., Robb, G. B. & Chan, S. H. mRNA capping: Biological functions and applications. *Nucleic Acids Research* **44**, 7511–7526 (2016).

[10] Sanford, J. R. Pre-mRNA splicing: life at the centre of the central dogma. *Journal of Cell Science* **117**, 6261–6263 (2004).

[11] Pandya-Jones, A. Pre-mRNA splicing during transcription in the mammalian system. *Wiley Interdisciplinary Reviews: RNA* **2**, 700–717 (2011).

[12] Neugebauer., K. M. On the importance of being co-transcriptional. *Journal of Cell Science* **115**, 3865–3871 (2002).

[13] Smith, C. W. J. & Valcárcel, J. Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends in Biochemical Sciences* **25**, 381–388 (2000).

[14] Wahl, M. C., Will, C. L. & Lührmann, R. The Spliceosome: Design Principles of a Dynamic RNP Machine. *Cell* **136**, 701–718 (2009).

[15] Will, C. L. & Max, R. L. Spliceosome Structure and Function. *Cold Spring Harb Perspect Biol* **3**, 2875–2881 (2002).

[16] Wang, Y. *et al.* Mechanism of alternative splicing and its regulation. *Biomedical Reports* **3**, 152–158 (2015).

[17] Martinez, N. M. & Lynch, K. W. Control of alternative splicing in immune responses: Many regulators, many predictions, much still to learn. *Immunol Rev.* **253**, 216–236 (2013).

[18] Chen, K., Dai, X. & Wu, J. Alternative splicing: An important mechanism in stem cell biology. *World journal of stem cells* **7**, 1–10 (2015).

[19] Keren, H., Lev-Maor, G. & Ast, G. Alternative splicing and evolution: Diversification, exon definition and function. *Nature Reviews Genetics* **11**, 345–355 (2010).

[20] Staiger, D. & Brown, J. W. S. Alternative Splicing at the Intersection of Biological Timing, Development, and Stress Responses. *The Plant Cell* **25**, 3640–3656 (2013).

[21] Roy, B. & Jacobson, A. The intimate relationships of mRNA decay and translation. *Trends in Genetics* **29**, 691–699 (2013).

[22] Vicente-Crespo, M. & Palacios, I. M. Nonsense-mediated mRNA decay and development: shoot the messenger to survive? *Biochemical Society transactions* **38**, 1500–5 (2010).

[23] Schoenberg, D. R. & Maquat, L. E. Regulation of cytoplasmic mrna decay. *Nature Reviews Genetics* **13**, 246–259 (2012).

[24] Nicholson, P. *et al.* Nonsense-mediated mRNA decay in human cells: Mechanistic insights, functions beyond quality control and the double-life of NMD factors. *Cellular and Molecular Life Sciences* **67**, 677–700 (2010).

[25] Garneau, N. L., Wilusz, J. & Wilusz, C. J. The highways and byways of mRNA decay. *Nature reviews. Molecular cell biology* **8**, 113–126 (2007).

[26] Shubhendu, G. & Jacobson, A. RNA decay modulates gene expression and controls its fidelity. *Wiley Interdiscip Rev RNA* **1**, 351–361 (2010).

[27] Kaikkonen, M. U., Lam, M. T. Y. & Glass, C. K. Non-coding RNAs as regulators of gene expression and epigenetics. *Cardiovascular Research* **90**, 430–440 (2011).

[28] Stroynowska-Czerwinska, A., Fiszer, A. & Krzyzosiak, W. J. The panorama of miRNA-mediated mechanisms in mammalian cells. *Cellular and Molecular Life Sciences* **71**, 2253–2270 (2014).

[29] Bartel, D. P. Micrornas: Target recognition and regulatory functions. *Cell* **136**, 215–233 (2009).

[30] Mortensen, R. D., Serra, M., Steitz, J. A. & Vasudevan, S. Posttranscriptional activation of gene expression in Xenopus laevis oocytes by microRNA-protein complexes (microRNPs). *Proceedings of the National Academy of Sciences* **108**, 8281–8286 (2011).

[31] Vasudevan, S., Tong, Y. & Steitz, J. A. Switching from repression to activation: Micrornas can up-regulate translation. *Science* **318**, 1931–1934 (2007).

[32] Kozomara, A. & Griffiths-Jones, S. Mirbase: Integrating microrna annotation and deep-sequencing data. *Nucleic Acids Research* **39**, 152–157 (2011).

[33] Hennig, J. & Sattler, M. Deciphering the protein-RNA recognition code: Combining large-scale quantitative methods with structural biology. *BioEssays* **37**, 899–908 (2015).

[34] Krol, J., Loedige, I. & Filipowicz, W. The widespread regulation of microRNA biogenesis, function and decay. *Nature Reviews Genetics* **11**, 597–610 (2010).

[35] Fabian, M. R., Sonenberg, N. & Filipowicz, W. Regulation of mRNA translation and stability by microRNAs. *Annual review of biochemistry* **79**, 351–379 (2010).

[36] Filipowicz, W., Bhattacharyya, S. N. & Sonenberg, N. Mechanisms of post-transcriptional regulation by microRNAs: Are the answers in sight? *Nature Reviews Genetics* **9**, 102–114 (2008).

[37] Treiber, T. *et al.* Resource A Compendium of RNA-Binding Proteins that Resource A Compendium of RNA-Binding Proteins that Regulate MicroRNA Biogenesis. *Molecular Cell* **66**, 270–281.e13 (2017).

[38] Finnegan, E. F. & Pasquinelli, A. E. MicroRNA biogenesis: regulating the regulators. *Critical reviews in biochemistry and molecular biology* **48**, 51–68 (2013).

[39] Cai, Y., Yu, X., Hu, S. & Yu, J. A Brief Review on the Mechanisms of miRNA Regulation. *Genomics, Proteomics and Bioinformatics* **7**, 147–154 (2009).

[40] Glisovic, T., Bachorik, J. L., Yong, J. & Dreyfuss, G. RNA-binding proteins and post-transcriptional gene regulation. *FEBS Letters* **582**, 1977–1986 (2008).

[41] Singh, G., Pratt, G., Yeo, G. W. & Moore, M. J. The Clothes Make the mRNA: Past and Present Trends in mRNP Fashion. *Annu Rev Biochem.* **84**, 325–354 (2015).

[42] Dominguez, D. *et al.* Sequence, Structure, and Context Preferences of Human RNA Binding Proteins. *Molecular Cell* **70**, 854–867.e9 (2018).

[43] Gerstberger, S., Hafner, M., Ascano, M. & Tuschl, T. Evolutionary Conservation and Expression of Human RNA-Binding Proteins and Their Role in Human Genetic Disease. In Yeo, E. W. (ed.) *Systems Biology of RNA Binding Proteins*, vol. 825, chap. 1, 1–55 (Springer, New York, 2014).

[44] Beckmann, B. M., Castello, A. & Medenbach, J. The expanding universe of ribonucleoproteins: of novel RNA-binding proteins and unconventional interactions. *Pflugers Archiv European Journal of Physiology* **468**, 1029–1040 (2016).

[45] Hentze, M. W., Castello, A., Schwarzl, T. & Preiss, T. A brave new world of RNA-binding proteins. *Nature Reviews Molecular Cell Biology* **19**, 327–341 (2018).

[46] Mayr, C. Regulation by 3'-untranslated regions. *Annu. Rev. Genet* **51**, 171–194 (2017).

[47] Cléry, A. & Frédéric, H.-T. A. From Structure to Function of RNA Binding Domains. In Lorković, Z. J. (ed.) *RNA Binding Proteins*, chap. 9, 137–158 (Landes Bioscience, 2012).

[48] Jolma, A. *et al.* Binding specificities of human rna binding proteins towards structured and linear rna sequences. *bioRxiv* (2018). `https://www.biorxiv.org/content/early/2018/05/16/317909.full.pdf`.

[49] Castello, A. *et al.* Comprehensive Identification of RNA-Binding Domains in Human Cells. *Molecular Cell* **63**, 696–710 (2016).

[50] Vinuesa, C. G. *et al.* A RING-type ubiquitin ligase family member required to repress follicular helper T cells and autoimmunity. *Nature* **435**, 452–458 (2005).

[51] Di Yu *et al.* Roquin represses autoimmunity by limiting inducible T-cell costimulator messenger RNA. *Nature* **450**, 299–303 (2007).

[52] Murakawa, Y. *et al.* RC3H1 post-transcriptionally regulates A20 mRNA and modulates the activity of the IKK/NF-κ B pathway. *Nature Communications* **6**, 7367 (2015).

[53] Vogel, K. U. *et al.* Roquin paralogs 1 and 2 redundantly repress the icos and ox40 costimulator mrnas and control follicular helper t cell differentiation. *Immunity* **38**, 655–668 (2013).

[54] Essig, K. *et al.* Roquin targets mRNAs in a 3'-UTR-specific manner by different modes of regulation. *Nature Communications* **9**, 3810 (2018).

[55] Bertossi, A. *et al.* Loss of Roquin induces early death and immune deregulation but not autoimmunity. *The Journal of Experimental Medicine* **208**, 1749–1756 (2011).

[56] Leppek, K. *et al.* Roquin promotes constitutive mrna decay via a conserved class of stem-loop recognition motifs. *Cell* **153**, 869–881 (2013).

[57] Schlundt, A., Niessing, D., Heissmeyer, V. & Sattler, M. RNA recognition by Roquin in posttranscriptional gene regulation. *Wiley Interdisciplinary Reviews: RNA* **7**, 455–469 (2016).

[58] Rehage, N. *et al.* Binding of NUFIP2 to Roquin promotes recognition and regulation of ICOS mRNA. *Nature Communications* **9**, 299 (2018).

[59] Schlundt, A. *et al.* Structural basis for RNA recognition in roquin-mediated post-transcriptional gene regulation. *Nature Structural & Molecular Biology* **21**, 671–678 (2014).

[60] Codutti, L. *et al.* A Distinct, Sequence-Induced Conformation Is Required for Recognition of the Constitutive Decay Element RNA by Roquin. *Structure* **23**, 1437–1447 (2015).

[61] Sakurai, S., Ohto, U. & Shimizu, T. Structure of human Roquin-2 and its complex with constitutive-decay element RNA. *Acta Crystallographica Section:F Structural Biology Communications* **71**, 1048–1054 (2015).

[62] Tan, D., Zhou, M., Kiledjian, M. & Tong, L. The ROQ domain of Roquin recognizes mRNA constitutive-decay element and double-stranded RNA. *Nature Publishing Group* **21**, 679–685 (2014).

[63] Zhang, Q. *et al.* New Insights into the RNA-Binding and E3 Ubiquitin Ligase Activities of Roquins. *Scientific Reports* **5**, 1–13 (2015).

[64] Janowski, R. *et al.* Roquin recognizes a non-canonical hexaloop structure in the 3´-UTR of Ox40. *Nature Communications* **7**, 11032 (2016).

[65] Essig, K. *et al.* Roquin Suppresses the PI3K-mTOR Signaling Pathway to Inhibit T Helper Cell Differentiation and Conversion of Treg to Tfr Cells. *Immunity* **47**, 1067–1082.e12 (2017).

[66] Li, L. *et al.* 3'UTR shortening identifies high-risk cancers with targeted dysregulation of the ceRNA network. *Scientific Reports* **4**, 5406 (2014).

[67] Ahmed, F., Benedito, V. A. & Zhao, P. X. Mining Functional Elements in Messenger RNAs: Overview, Challenges, and Perspectives. *Frontiers in plant science* **2**, 84 (2011).

[68] Mignone, F., Gissi, C., Liuni, S. & Pesole, G. Untranslated regions of mRNAs. *Genome Biology* **3**, REVIEWS0004 (2002).

[69] Araujo, P. R. *et al.* Before it gets started: Regulating translation at the 5'utr. *Comparative and Functional Genomics* **2012**, 475731 (2012).

[70] Huin, V. *et al.* Alternative promoter usage generates novel shorter MAPT mRNA transcripts in Alzheimer's disease and progressive supranuclear palsy brains. *Scientific Reports* **7**, 12589 (2017).

[71] Barbosa, C., Peixeiro, I. & Romão, L. Gene Expression Regulation by Upstream Open Reading Frames and Human Disease. *PLoS Genetics* **9**, e1003529 (2013).

[72] Kunze, M. M. *et al.* sST2 translation is regulated by FGF2 via an hnRNP A1-mediated IRES-dependent mechanism. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1859**, 848–859 (2016).

[73] Bonnal, S. IRESdb: the Internal Ribosome Entry Site database. *Nucleic Acids Research* **31**, 427–428 (2003).

[74] Hellen, C. U. T. & Sarnow, P. Internal ribosome entry sites in eukaryotic mRNA molecules. *Genes and Development* **15**, 1593–1612 (2001).

[75] Martínez-Salas, E., Francisco-Velilla, R., Fernandez-Chamorro, J., Lozano, G. & Diaz-Toledano, R. Picornavirus IRES elements: RNA structure and host protein interactions. *Virus Research* **206**, 62–73 (2015).

[76] Van Eden, M. E., Byrd, M. P., Sherrill, K. W. & Lloyd, R. E. Translation of cellular inhibitor of apoptosis protein 1 (c-IAP1) mRNA is IRES mediated and regulated during cell stress. *RNA* **10**, 469–481 (2004).

[77] Vagner, S. *et al.* Alternative translation of human fibroblast growth factor 2 mRNA occurs by internal entry of ribosomes. *Molecular and Cellular Biology* **15**, 35–44 (1995).

[78] Lang, K. J. D., Kappel, A., Goodall, G. J. & Wickens, M. P. Hypoxia-inducible factor-1α mrna contains an internal ribosome entry site that allows efficient translation during normoxia and hypoxia. *Molecular Biology of the Cell* **13**, 1792–1801 (2002).

[79] Hentze, M. W. *et al.* Identification of the iron-responsive element for the translational regulation of human ferritin mRNA. *Science* **238**, 1570–1573 (1987).

[80] Hentze, M. W. & Kuhnt, L. C. Molecular control of vertebrate iron metabolism: mRNA-based regulatory circuits operated by iron, nitric oxide, and oxidative stress. *Proc. Natl. Acad. Sci. USA* **93**, 8175–8182 (1996).

[81] Subbarao, K. *et al.* Structure of Dual Function Iron Regulatory Protein 1 Complexed with Ferritin IRE-RNA. *Science* **314**, 1903–1909 (2006).

[82] Bakheet, T., Hitti, E., Al-Saif, M., Moghrabi, W. N. & Khabar, K. S. The AU-rich element landscape across human transcriptome reveals a large proportion in introns and regulation by ELAVL1/HuR. *BBA - Gene Regulatory Mechanisms* **1861**, 167–177 (2017).

[83] Barreau, C., Paillard, L. & Osborne, H. B. AU-rich elements and associated factors: Are there unifying principles? *Nucleic Acids Research* **33**, 7138–7150 (2005).

[84] Schiavi, S. C., Belasco, J. G. & Greenberg, M. E. Regulation of proto-oncogene mRNA stability. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer* **1114**, 95–106 (1992).

[85] Khabar, K. S. Hallmarks of cancer and AU-rich elements. *Wiley Interdisciplinary Reviews: RNA* **8**, e1368 (2017).

[86] De Toeuf, B. *et al.* ARE-mediated decay controls gene expression and cellular metabolism upon oxygen variations. *Scientific Reports* **8**, 5211 (2018).

[87] Halees, A. S. *et al.* Global assessment of gu-rich regulatory content and function in the human transcriptome. *RNA Biology* **8**, 681–691 (2011).

[88] Vlasova, I. A. & Bohjanen, P. R. Posttranscriptional regulation of gene networks by GU-rich elements and CELF proteins. *RNA Biol. 2008* **5**, 201–207 (2008).

[89] Vlasova, I. A. *et al.* Conserved GU-rich elements mediate mRNA decay by binding to CUG-binding protein 1. *Molecular cell* **29**, 263–70 (2008).

[90] Cook, K. B., Kazan, H., Zuberi, K., Morris, Q. & Hughes, T. R. RBPDB: a database of RNA-binding specificities. *Nucleic Acids Research* **39**, D301–D308 (2011).

[91] Wong, N. & Wang, X. mirdb: an online resource for microrna target prediction and functional annotations. *Nucleic Acids Research* **43**, D146–D152 (2015).

[92] Bakheet, T., Hitti, E. & Khabar, K. Ared-plus: an updated and expanded database of au-rich element-containing mrnas and pre-mrnas. *Nucleic Acids Research* **46**, D218–D220 (2018).

[93] Piva, F., Giulietti, M., Nocchi, L. & Principato, G. Spliceaid: a database of experimental rna target motifs bound by splicing proteins in humans. *Bioinformatics* **25**, 1211–1213 (2009).

[94] Parker, B. J. *et al.* New families of human regulatory RNA structures identified by comparative analysis of vertebrate genomes. *Genome Research* **21**, 1929–1943 (2011).

[95] Batey, R. T., Rambo, R. P. & Doudna, J. A. Tertiary Motifs in RNA Structure and Folding Robert. *Angew. Chem. Int. Ed.* **38**, 2326–2343 (1999).

[96] Marzluff, W. F., Wagner, E. J. & Duronio, R. J. Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. *Nat Rev Genet.* **9**, 843–854 (2008).

[97] Mullen, T. E. & Marzluff, W. F. Degradation of histone mRNA requires oligouridylation followed by decapping and simultaneous degradation of the mRNA both 5' to 3' and 3' to 5'. *Genes and Development* **22**, 50–65 (2008).

[98] Su, W. *et al.* mRNAs containing the histone 3' stem-loop are degraded primarily by decapping mediated by oligouridylation of the 3' end. *RNA* **19**, 1–16 (2013).

[99] Hoefig, K. P. *et al.* Eri1 degrades the stem-loop of oligouridylated histone mRNAs to induce replication-dependent decay. *Nature Structural and Molecular Biology* **20**, 73–81 (2013).

[100] Aviv, T. *et al.* The RNA-binding SAM domain of Smaug defines a new family of post-transcriptional regulators. *Nature Structural Biology* **10**, 614–621 (2003).

[101] Yoichiro, S. *et al.* hiCLIP reveals the in vivo atlas of mRNA secondary structures recognized by Staufen 1. *Nature* **7544**, 491–494 (2015).

[102] Park, E. & Maquat, L. E. Staufen-mediated mrna decay. *Wiley Interdisciplinary Reviews: RNA* **4**, 423–435 (2013).

[103] Gong, C. & Maquat, L. E. lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements. *Nature* **470**, 284–288 (2011).

[104] Latrèche, L., Jean-Jean, O., Driscoll, D. M. & Chavatte, L. Novel structural determinants in human SECIS elements modulate the translational recoding of UGA as selenocysteine. *Nucleic Acids Research* **37**, 5868–5880 (2009).

[105] Lee, F. C. Y. & Ule, J. Advances in CLIP Technologies for Studies of Protein-RNA Interactions. *Molecular cell* **69**, 354–369 (2018).

[106] Wheeler, E. C., Van Nostrand, E. L. & Yeo, G. W. Advances and challenges in the detection of transcriptome-wide protein–rna interactions. *Wiley Interdisciplinary Reviews: RNA* **9**, e1436 (2018).

[107] Wissink, E. M., Fogarty, E. A. & Grimson, A. High-throughput discovery of post-transcriptional cis-regulatory elements. *BMC Genomics* **17**, 177 (2016).

[108] Flynn, R. A. *et al.* Transcriptome-wide interrogation of RNA secondary structure in living cells with icSHAPE. *Nature Protocols* **11**, 273–290 (2016).

[109] Rouskin, S., Zubradt, M., Washietl, S., Kellis, M. & Weissman, J. S. Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature* **505**, 701–705 (2013).

[110] Spitale, R. C. *et al.* RNA SHAPE analysis in living cells. *Nature Chemical Biology* **9**, 18–20 (2012).

[111] Talkish, J., May, G., Lin, Y., Woolford, J. L. & McManus, C. J. Mod-seq: High-throughput sequencing for chemical probing of RNA structure. *Rna* **20**, 713–720 (2014).

[112] Lu, Z. *et al.* RNA Duplex Map in Living Cells Reveals Higher-Order Transcriptome Structure. *Cell* **165**, 1267–1279 (2016).

[113] Braun, J. *et al.* Identification of new high affinity targets for Roquin based on structural conservation. *Nucleic Acids Research* gky908 (2018). `https://doi.org/10.1093/nar/gky908`.

[114] Hofacker, I. L. Energy-directed rna structure prediction. In Gorodkin, J. & Ruzzo, W. L. (eds.) *RNA Sequence, Structure, and Function: Computational and Bioinformatic Methods*, 71–84 (Humana Press, Totowa, NJ, 2014).

[115] Mathews, D. H. & Turner, D. H. Prediction of rna secondary structure by free energy minimization. *Current Opinion in Structural Biology* **16**, 270 – 278 (2006). Nucleic acids/Sequences and topology.

[116] Mathews, D. H. *et al.* Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 7287–7292 (2004).

[117] McCaskill, J. S. The equilibrium partition function and base pair binding probabilities for rna secondary structure. *Biopolymers* **29**, 1105–1119 (1990).

[118] Pace, N., Thomas, B. & Woese, C. Probing rna structure, function, and history by comparative analysis. In Gesteland, R. F., Cech, T. R. & Atkins, J. (eds.) *The RNA World: The Nature of Modern RNA Suggests a Prebiotic RNA World*, vol. 37, chap. 4, 113–141 (Cold Spring Harbor Laboratory Press, 1999), 2 edn.

[119] Rivas, E. & Eddy, S. R. Noncoding RNA gene detection using comparative sequence analysis. *BMC Bioinformatics* **2**, 1–19 (2001).

[120] Parsch, J., Braverman, J. M. & Stephan, W. Comparative sequence analysis and patterns of covariation in RNA secondary structures. *Genetics* **154**, 909–921 (2000).

[121] Weinberg, Z. *et al.* Comparative genomics reveals 104 candidate structured RNAs from bacteria, archaea, and their metagenomes. *Genome Biology* **11**, R31 (2010).

[122] Washietl, S., Hofacker, I. L. & Stadler, P. F. From The Cover: Fast and reliable prediction of noncoding RNAs. *Proceedings of the National Academy of Sciences* **102**, 2454–2459 (2005).

[123] Gruber, A. R., Findeiss, S., Washietl, S., Hofacker, I. L. & Stadler, P. F. Rnaz 2.0: Improved noncoding rna detection. *Pac Symp Biocomput.* **15**, 69–79 (2010).

[124] Torarinsson, E., Sawera, M., Havgaard, J. H., Fredholm, M. & Gorodkin, J. Thousands of corresponding human and mouse genomic regions unalignable in primary sequence contain common RNA structure. *Genome Research* **16**, 885–889 (2006).

[125] Uzilov, A. V., Keegan, J. M. & Mathews, D. H. Detection of non-coding RNAs on the basis of predicted secondary structure formation free energy change. *BMC Bioinformatics* **7**, 173 (2006).

[126] Havgaard, J. H., Lyngsø, R. B. & Gorodkin, J. The Foldalign web server for pairwise structural RNA alignment and mutual motif search. *Nucleic Acids Research* **33**, W650–W653 (2005).

[127] Mathews, D. H. & Turner, D. H. Dynalign: an algorithm for finding the secondary structure common to two rna sequences11edited by i. tinoco. *Journal of Molecular Biology* **317**, 191 – 203 (2002).

[128] Bouillaud, F., Alves-Guerra, M. C. & Ricquier, D. UCPs, at the interface between bioenergetics and metabolism. *Biochimica et Biophysica Acta - Molecular Cell Research* **1863**, 2443–2464 (2016).

[129] Busiello, R. A., Savarese, S. & Lombardi, A. Mitochondrial uncoupling proteins and energy metabolism. *Frontiers in Physiology* **6**, 36 (2015).

[130] Shepard, P. J. & Hertel, K. J. Conserved RNA secondary structures promote alternative splicing. *RNA* **14**, 1463–1469 (2008).

[131] Stefanie A., M., Mary Anne, K. & Jennifer A., D. Insights into RNA structure and function from genome-wide studies . *Nature Reviews Genetics* **15**, 469–479 (2014).

[132] Mathews, D. H. & Turner, D. H. Dynalign: an algorithm for finding the secondary structure common to two RNA sequences. *Journal of molecular biology* **317**, 191–203 (2002).

[133] Harmanci, A. O., Sharma, G. & Mathews, D. H. Efficient pairwise RNA structure prediction using probabilistic alignment constraints in Dynalign. *BMC Bioinformatics* **8**, 130 (2007).

[134] Havgaard, J., J. H. Gorodkin. RNA structural alignments, part I: Sankoff-based approaches for structural alignments. *Methods Mol Biol* **1097**, 275–290 (2014).

[135] Lu, Z. J. *Secondary structure prediction of non-coding RNA*. Ph.D. thesis, University of Rochester School of Medicine and Dentistry, Rochester, N.Y. (2008).

[136] Jezek, P., Holendová, B., Garlid, K. D. & Jaburek, M. Mitochondrial uncoupling proteins: subtle regulators of cellular redox signaling. *Antioxidants & Redox Signaling* **29**, 667–714 (2018).

[137] Hilse, K. E. *et al.* The expression of uncoupling protein 3 coincides with the fatty acid oxidation type of metabolism in adult murine heart. *Frontiers in Physiology* **9**, 747 (2018).

[138] Agarwal, V., Bell, G. W., Nam, J. W. & Bartel, D. P. Predicting effective microRNA target sites in mammalian mRNAs. *eLife* **4**, e05005 (2015).

[139] Betel, D., Koppal, A., Agius, P., Sander, C. & Leslie, C. Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. *Genome Biology* **11** (2010).

[140] Varkonyi-Gasic E., H. R. Quantitative Stem-Loop RT-PCR for Detection of MicroRNAs? *Methods in Molecular Biology* **744**, 102–114 (2011).

[141] Roux, M. *et al.* Plasma levels of hsa-mir-152-3p are associated with diabetic nephropathy in patients with type 2 diabetes. *Nephrology Dialysis Transplantation* gfx367 (2018). `https://doi.org/10.1093/ndt/gfx367`.

[142] Ofori, J. K. *et al.* Elevated miR-130a/miR130b/miR-152 expression reduces intracellular ATP levels in the pancreatic beta cell. *Scientific Reports* **7**, 44986 (2017).

[143] Chen, C.-Y. A., Ezzeddine, N. & Shyu, A.-B. Messenger RNA Half-Life Measurements in Mammalian Cells. *Methods Enzymol* **448**, 335–357 (2008).

[144] Thakor, N. & Holcik, M. Rna affinity chromatography. In Magdeldin, S. (ed.) *Affinity Chromatography*, chap. 8 (IntechOpen, Rijeka, 2012).

[145] Glasmacher, E. *et al.* Roquin binds inducible costimulator mRNA and effectors of mRNA decay to induce microRNA- independent post-transcriptional repression. *Nature Immunology* **11**, 725–733 (2011).

[146] Charles, A., Paul G., C., Nadeene, P. & Martin, D. B. Novel uncoupling proteins. In Derek J., C. & Jamie, G. (eds.) *MITOCHONDRIAL BIOLOGY: NEW PERSPECTIVES*, chap. 5, 70–92 (John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester PO19 8SQ, UK, 2007), 1 edn.

[147] Azzu, V., Jastroch, M., Divakaruni, A. S. & Brand, M. D. The regulation and turnover of mitochondrial uncoupling proteins. *Biochimica et Biophysica Acta - Bioenergetics* **1797**, 785–791 (2010).

[148] Azzu, V. & Brand, M. D. The on-off switches of the mitochondrial uncoupling proteins. *Trends in Biochemical Sciences* **35**, 298–307 (2010).

[149] Ryder, S. P., Recht, M. I. & Williamson, J. R. Quantitative analysis of protein-RNA interactions by gel mobility shift. *Methods Mol Biol.* **488**, 99–115 (2008).

[150] Lu, Z. J. & Mathews, D. H. Fundamental differences in the equilibrium considerations for siRNA and antisense oligodeoxynucleotide design. *Nucleic Acids Research* **36**, 3738–3745 (2008).

[151] Sloma, M. F. & Mathews, D. H. Base pair probability estimates improve the prediction accuracy of RNA non-canonical base pairs. *PLoS Computational Biology* **13**, e1005827 (2017).

[152] Day, D. A. & Tuite, M. F. Post-transcriptional gene regulatory mechanisms in eukaryotes: an overview. *The Journal of endocrinology* **157**, 361–371 (1998).

[153] Robinson, S., Oliver, A., Chevassut, T. & Newbury, S. The 3' to 5' Exoribonuclease DIS3: From Structure and Mechanisms to Biological Functions and Role in Human Disease. *Biomolecules* **5**, 1515–1539 (2015).

[154] Sharp, P. A. The Centrality of RNA. *Cell* **136**, 577–580 (2009).

[155] Westhof, E. Twenty Years of RNA crystallography. *RNA* **21**, 486–487 (2015).

[156] Boris, F., Christian, R., Jens, W. & Harald, S. NMR Spectroscopy of RNA. *ChemBioChem* **4**, 936–962 (2003).

[157] Ledda, M. & Aviran, S. PATTERNA: Transcriptome-wide search for functional RNA elements via structural data signatures. *Genome Biology* **19**, 28 (2018). `https://doi.org/10.1186/s13059-018-1399-z`.

[158] Choudhary, K., Deng, F. & Aviran, S. Comparative and integrative analysis of RNA structural profiling data: current practices and emerging questions. *Quantitative Biology* **5**, 3–24 (2017).

[159] Spitale, R. C. *et al.* RNA SHAPE analysis in living cells. *Nat Chem Biol* **9**, 18–20 (2013).

[160] Baltz, A. G. *et al.* The mRNA-Bound Proteome and Its Global Occupancy Profile on Protein-Coding Transcripts. *Molecular Cell* **46**, 674–690 (2012).

[161] Castello, A. *et al.* Insights into RNA Biology from an Atlas of Mammalian mRNA-Binding Proteins. *Cell* **149**, 1393–1406 (2012).

[162] Mathews, D. H. & Turner, D. H. Prediction of RNA secondary structure by free energy minimization. *Current Opinion in Structural Biology* **16**, 270–278 (2006).

[163] Pedersen, J. S. *et al.* Identification and classification of conserved RNA secondary structures in the human genome. *PLoS Computational Biology* **2**, 251–262 (2006).

[164] Spieth, J., Hillier, L. W. & Wilson, R. K. Evolutionarily conserved elements in vertebrate , insect , worm , and yeast genomes. *Nature* **15**, 1034–1050 (2005).

[165] Xie, X. *et al.* Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. *Nature* **434**, 338–345 (2005).

[166] Xu, Z. & Mathews, D. H. Multilign: An algorithm to predict secondary structures conserved in multiple RNA sequences. *Bioinformatics* **27**, 626–632 (2011).

[167] Biggar, K. K. & Storey, K. B. Insight into post-transcriptional gene regulation: stress-responsive microRNAs and their role in the environmental stress survival of tolerant animals. *Journal of Experimental Biology* **218**, 1281–1289 (2015).

[168] Thomas, M. P. & Lieberman, J. Live or let die: posttranscriptional gene regulation in cell stress and cell death. *Immunological Reviews* **253**, 237–252 (2013).

[169] Wenxue, Z. *et al.* Massively parallel functional annotation of 3' untranslated regions Wenxue. *Nat Biotechnol* **32**, 387–391 (2014).

[170] Oikonomou, P., Goodarzi, H. & Tavazoie, S. Systematic identification of regulatory elements in conserved 3' UTRs of human transcripts. *Cell Reports* **7**, 281–292 (2014).

[171] Schmitter, D. *et al.* Effects of Dicer and Argonaute down-regulation on mRNA levels in human HEK293 cells. *Nucleic Acids Research* **34**, 4801–4815 (2006).

[172] Yoon, J. H. *et al.* PAR-CLIP analysis uncovers AUF1 impact on target RNA fate and genome integrity. *Nature Communications* **5**, 5248 (2014).

[173] Qi, M.-Y. *et al.* AU-Rich-Element-Dependent Translation Repression Requires the Cooperation of Tristetraprolin and RCK/P54. *Molecular and Cellular Biology* **32**, 913–928 (2012).

[174] Renold, A., Koehler, C. M. & Murphy, M. P. Mitochondrial import of the long and short isoforms of human uncoupling protein 3. *FEBS Letters* **465**, 135–140 (2000).

[175] Krauss, S., Zhang, C.-y. & Lowell, B. B. THE MITOCHONDRIAL HOMOLOGUES. *Nat Rev Mol Cell Biol.* **6**, 248–261 (2005).

[176] Brand, M. D. & Esteves, T. C. Physiological functions of the mitochondrial uncoupling proteins UCP2 and UCP3. *Cell Metabolism* **2**, 85–93 (2005).

[177] Bézaire, V., Seifert, E. L. & Harper, M.-E. Uncoupling protein-3: clues in an ongoing mitochondrial mystery. *The FASEB Journal* **21**, 312–324 (2007).

[178] Gong, D. W. *et al.* Lack of obesity and normal response to fasting and thyroid hormone in mice lacking uncoupling protein-3. *Journal of Biological Chemistry* **275**, 16251–16257 (2000).

[179] Vidal-Puig, A. J. *et al.* Energy metabolism in uncoupling protein 3 gene knockout mice. *Journal of Biological Chemistry* **275**, 16258–16266 (2000).

[180] Hesselink, M. K., Mensink, M. & Schrauwen, P. Human uncoupling protein-3 and obesity: An update. *Obesity Research* **11**, 1429–1443 (2003).

[181] Schrauwen, P. *et al.* Reduced skeletal muscle uncoupling protein-3 content in prediabetic subjects and type 2 diabetic patients: Restoration by rosiglitazone treatment. *Journal of Clinical Endocrinology and Metabolism* **91**, 1520–1525 (2006).

[182] Hwang, C.-S. & Lane, M. Up-Regulation of Uncoupling Protein-3 by Fatty Acid in C2C12 Myotubes. *Biochemical and Biophysical Research Communications* **258**, 464–469 (1999).

[183] Friedman, R. C., Farh, K. K.-h., Burge, C. B. & Bartel, D. P. Most mammalian mRNAs are conserved targets of miRNAs. *Genome Research* **19**, 92–105 (2009).

[184] Liu, X., Li, J., Qin, F. & Dai, S. MiR-152 as a tumor suppressor microRNA: Target recognition and regulation in cancer. *Oncology Letters* **11**, 3911–3916 (2016).

[185] Kathrin Leppek. *RNA Affinity Purification and Characterization of Roquin Proteins in CDE-mediated mRNA Decay*. Ph.D. thesis, Ruperto-Carola University of Heidelberg, Germany (2014).

[186] Schuetz, A., Murakawa, Y., Rosenbaum, E., Landthaler, M. & Heinemann, U. Roquin binding to target mRNAs involves a winged helix-turn-helix motif. *Nature Communications* **5**, 5701 (2014).

[187] Heissmeyer, V. & Vogel, K. U. Molecular control of tfh-cell differentiation by roquin family proteins. *Immunological Reviews* **253**, 273–289 (2013).

[188] Hall, T. M. T. Multiple modes of RNA recognition by zinc finger proteins. *Current Opinion in Structural Biology* **15**, 367–373 (2005).

[189] Fu, M. & J., B. P. RNA-binding proteins in immune regulation: a focus on CCCH zinc finger proteins. *Nature Reviews Immunology* **8**, 444–454 (2016).

[190] Mino, T. *et al.* Regnase-1 and roquin regulate a common element in inflammatory mRNAs by spatiotemporally distinct mechanisms. *Cell* **161**, 1058–1073 (2015).

[191] Jeltsch, K. M. *et al.* Cleavage of roquin and regnase-1 by the paracaspase MALT1 releases their cooperatively repressed targets to promote TH17 differentiation. *Nature Immunology* **15**, 1079–1089 (2014).

[192] Yoshinaga, M. *et al.* Regnase-1 Maintains Iron Homeostasis via the Degradation of Transferrin Receptor 1 and Prolyl-Hydroxylase-Domain-Containing Protein 3 mRNAs. *Cell Reports* **19**, 1614–1630 (2017).

[193] Cui, X. *et al.* Regnase-1 and roquin nonredundantly regulate th1 differentiation causing cardiac inflammation and fibrosis. *The Journal of Immunology* **199**, 4066–4077 (2017).

[194] Iwasaki, H. *et al.* The IκB kinase complex regulates the stability of cytokine-encoding mRNA induced by TLR–IL-1R by controlling degradation of regnase-1. *Nature Immunology* **12**, 1167 (2011).

[195] Putland, R. A. *et al.* RNA Destabilization by the Granulocyte Colony-Stimulating Factor Stem-Loop Destabilizing Element Involves a Single Stem-Loop That Promotes Deadenylation. *Molecular and Cellular Biology* **22**, 1664–1673 (2002).

[196] Hattori, K., Naguro, I., Runchel, C. & Ichijo, H. The roles of ASK family proteins in stress responses and diseases. *Cell Communication and Signaling* **7**, 9 (2009).

[197] Maruyama, T. *et al.* Roquin-2 promotes ubiquitin-mediated degradation of ask1 to regulate stress responses. *Science Signaling* **7**, pp.ra8 (2014).

[198] Hanieh, H. *et al.* Arid5a stabilizes ox40 mrna in murine cd4+ t cells by recognizing a stem-loop structure in its 3´utr. *European Journal of Immunology* **48**, 593–604 (2018).

[199] An, M. X. *et al.* BAG3 directly stabilizes Hexokinase 2 mRNA and promotes aerobic glycolysis in pancreatic cancer cells. *Journal of Cell Biology* **216**, 4091–4105 (2017).

[200] Carballo, E., Lai, W. S. & Blackshear, P. J. Feedback Inhibition of Macrophage Tumor Necrosis Factor-α Production by Tristetraprolin. *Science* **281**, 1001–1005 (1998).

[201] Srivastava, M. *et al.* Roquin binds microRNA-146a and Argonaute2 to regulate microRNA homeostasis. *Nature Communications* **6**, 6253 (2015).

[202] Kemmerer, K. & Weigand, J. E. Hypoxia reduces max expression in endothelial cells by unproductive splicing. *FEBS Letters* **588**, 4784–4790 (2014).

[203] Higuchi, R., Krummel, B. & Saiki, R. A general method of in vitro preparation and specific mutagenesis of dna fragments: Study of protein and DNA interactions. *Nucleic Acids Research* **16**, 7351–7367 (1988).

[204] Walker, S. C. General plasmids for producing RNA in vitro transcripts with homogeneous ends. *Nucleic Acids Research* **31**, e82 (2003).

[205] Regulski, E. E. & Breaker, R. R. In-line probing analysis of riboswitches. *Methods in Molecular Biology* **419**, 53–67 (2008).

[206] Mathews, D. H. Predicting a set of minimal free energy rna secondary structures common to two sequences. *Bioinformatics* **21**, 2246–2253 (2005).

[207] Macke, T., Ecker, D. & Gutell, R. RNAMotif, an RNA secondary structure definition and search algorithm. *Nucleic Acids Research* **29**, 4724–4735 (2001).

[208] Reuter, J. S. *et al.* RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* **11**, 129 (2010).

[209] Karolchik, D. *et al.* The ucsc table browser data retrieval tool. *Nucleic Acids Research* **32**, D493–D496 (2004).

[210] Blanchette, M. *et al.* Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Research* **14**, 708–715 (2004).

Ehrenwörtliche Erklärung

Ich erkläre hiermit ehrenwörtlich, dass ich die vorliegende Arbeit entsprechend den Regeln guter wissenschaftlicher Praxis selbstständig und ohne unzulässige Hilfe Dritter angefertigt habe.

Sämtliche aus fremden Quellen direkt oder indirekt übernommenen Gedanken sowie sämtliche von Anderen direkt oder indirekt übernommenen Daten, Techniken und Materialien sind als solche kenntlich gemacht. Die Arbeit wurde bisher bei keiner anderen Hochschule zu Prüfungszwecken eingereicht.

Darmstadt den 04.12.2018

Johannes Braun

Curriculum Vitae

Personal Information

Name:	Johannes Braun
Date of birth	26.01.1986
Address:	Dilsbachstr. 21, 64354 Reinheim

Experience

08/2014–10/2018	**Research Associate: RNA Biology**, TU Darmstadt, Germany
04/2012–09/2012	**Research Engineer: BioActives and Performance Biologicals**, Brain AG, Germany
09/2011–02/2012	**Intern: Producer Strain Development**, Brain AG, Germany

Education

10/2012–07/2014	**Technical University Darmstadt**, Germany **graduation**: 08/2014, *M. Sc. Technical Biology*
10/2007–03/2012	**University of Applied Sciences Darmstadt**, Germany **graduation**: 03/2012, *B. Sc. Biotechnology*
10/2006–03/2007	**Maximiliansuniversität Würzburg**, Germany
1996–2006	**Otto-Hahn-Gymnasium Landau, i. d. Pfalz**, Germany

Talks and Poster Presentations

2018 **Understanding RNA-based Regulation in Cells**, Frankfurt, Germany
Talk: *Identification of high affinity Roquin targets based on structural conservation*

2018 **RNA Structural Biology**, Bad Homburg, Germany
Talk: *Mining for new cis-regulatory elements in messenger RNAs*

2017 **35. Rabensteiner Kolleg**, Pottenstein, Germany
Talk: *Mining for new cis-regulatory elements in messenger RNAs*

2016 **34. Rabensteiner Kolleg**, Pottenstein, Germany
Talk: *Postranscriptional regulation of UCP3*

2017 **RNA Society: RNA 2017**, Prague, Czech Republic
Poster: *Mining for new cis-regulatory elements in messenger RNAs*

2016 **Analysis of NextGen RNA-Seq data for expression profiling and protein-binding RNAs**, Regensburg, Germany
Poster: *Mining for new cis-regulatory elements in messenger RNAs*

2015 **EMBO: Protein synthesis and translational control**, Heidelberg, Germany
Poster: *Mining for new cis-regulatory elements in messenger RNAs*

5 Supplemental Data

5.1 Supplementary Figures

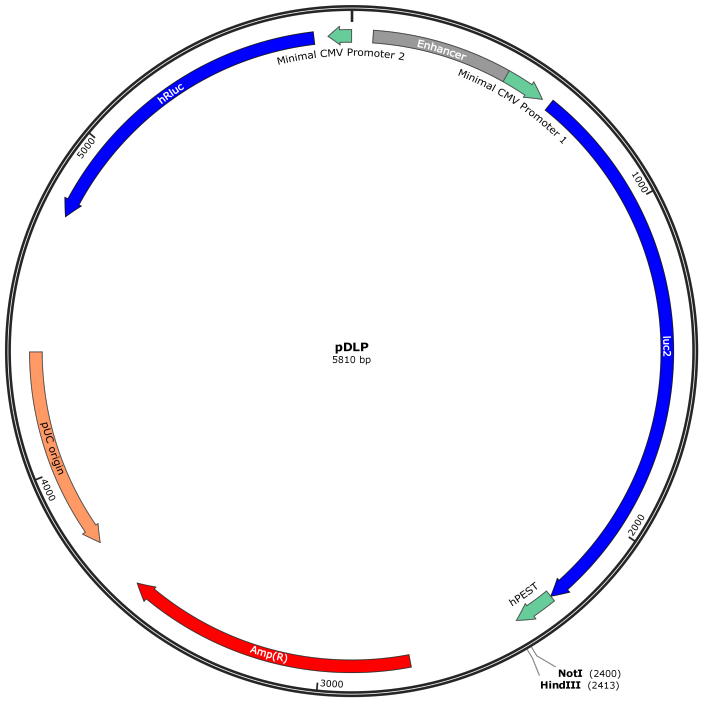

Figure 5.1: Schematic map of the pDLP plasmid. The plasmid contains the CDS of the β-lactamase (AmpR), firefly luciferase (luc2) and *Renilla* luciferase (hRluc) expressed from a CMV promoter (Minimal CMV Promoter 1/2) and a shared enhancer element (Enhancer). Additional features:, origin of replication (pUC origin), destabilization sequence from the mouse ornithine decarboxylase (hPEST). Recognition sites for the restriction endonuleases NotI and HindIII are indicated.

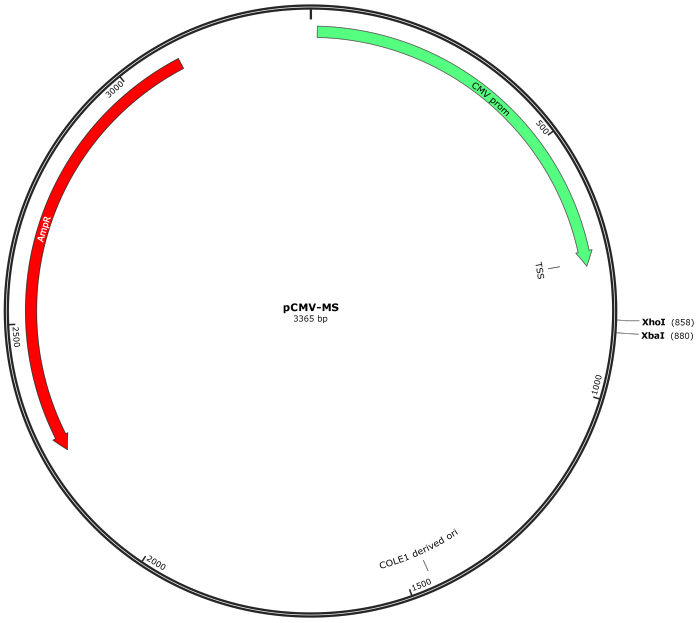

Figure 5.2: Schematic map of the pCMV-MS plasmid. The plasmid contains the CDS of the β-lactamase (AmpR) for ampillicin resistance and a CMV promoter for expression of downstream sequences. Recognition sites for the restriction endonuleases XhoI and XbaI are indicated.

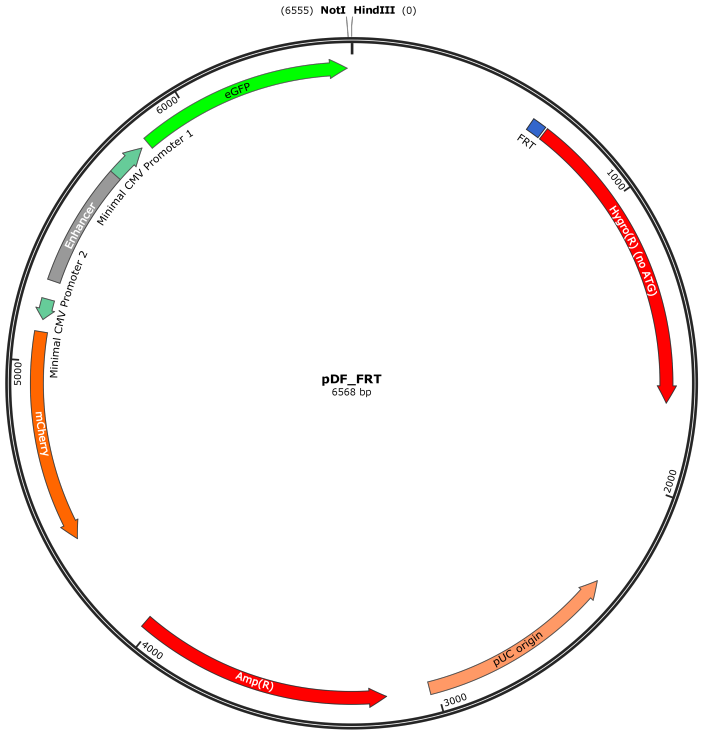

Figure 5.3: Schematic map of the pDF_FRT plasmid. The plasmid contains the CDS of the β-lactamase (AmpR), hygromycin resistance gene (Hygro(R)), enhanced GFP (eGFP) and mCHerry expressed from a CMV promoter (Minimal CMV Promoter 1/2) and a shared enhancer element (Enhancer). Additional features: origin of replication (pUC origin). Recognition sites for the restriction endonuleases NotI and HindIII are indicated.

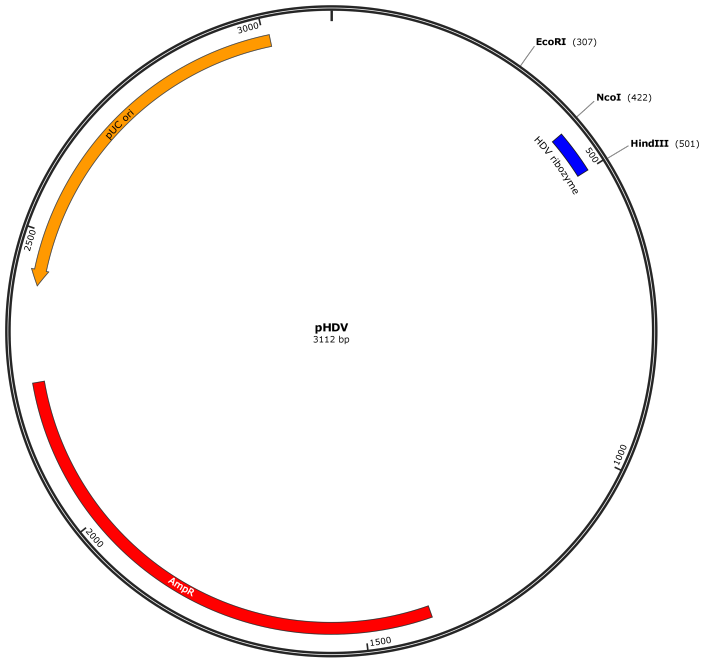

Figure 5.4: Schematic map of the pHDV plasmid. The plasmid contains the CDS of the β-lactamase (AmpR) for ampillicin resistance and the sequence of the HDV ribozym. Additional features: origin of replication (pUC origin). Recognition sites for the restriction endonuleases EcoRI, NcoI and HindIII are indicated.

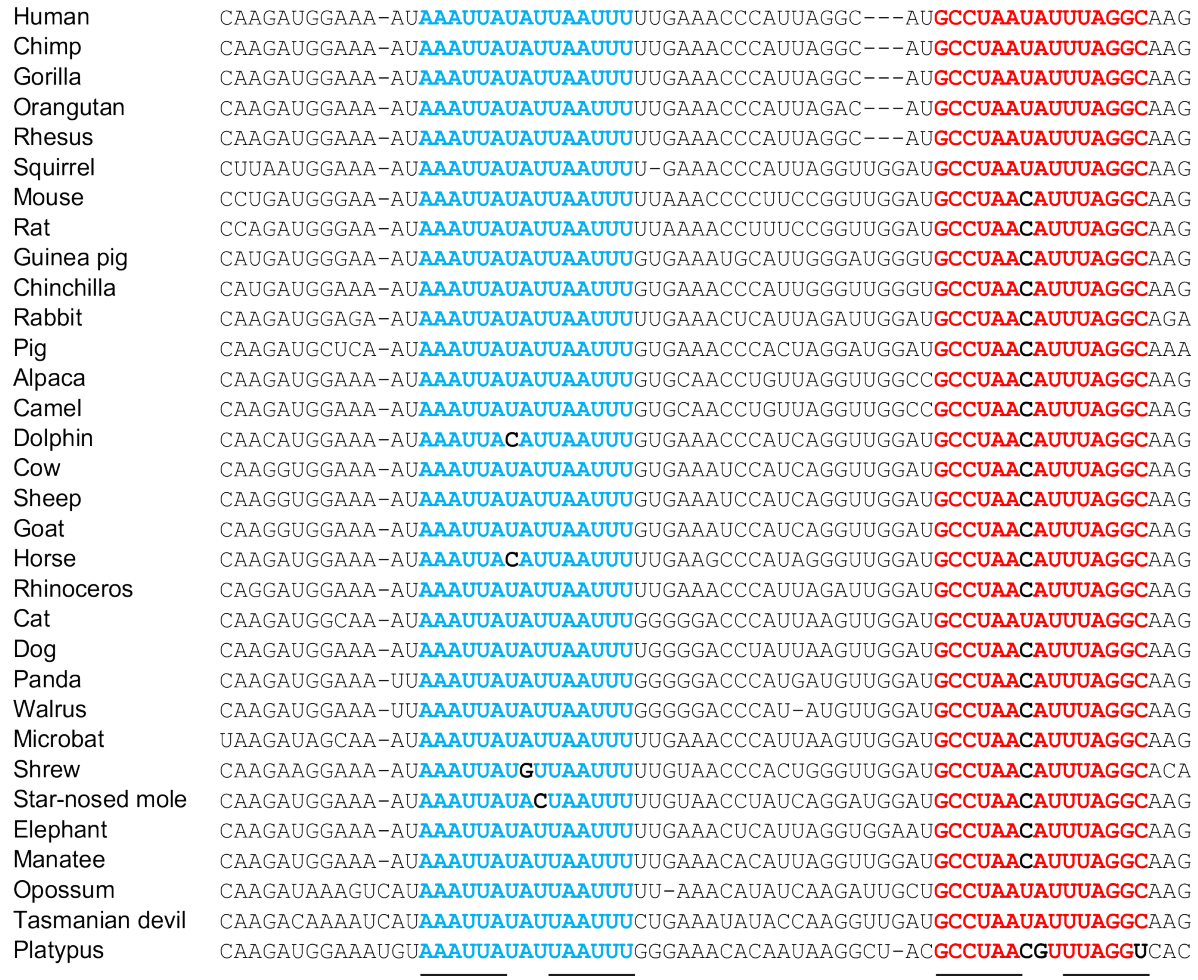

Figure 5.5: Conservation of the CDEs in UCP3 across mammals. CDEI and CDEII are highlighted in blue and red, respectively. Mutations within CDEs relative to the human sequence are indicated in black/bold. Adopted and modified from [113].

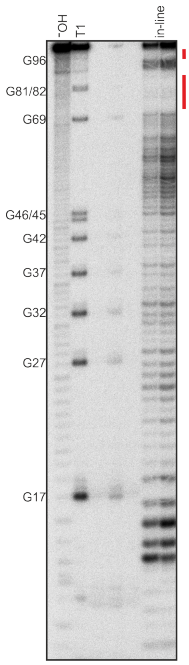

Figure 5.6: In-line probing of the predicted structurally conserved *UCP3* element. In-line probing analysis of the 100 nt stucurally conserved elements in the UCP3 3'UTR. The RNA was subjected to cleavage by RNase T1 or alkaline hydrolysis, or incubated for the indicated time points at room temperature and pH 8.3 (in-line) prior to Urea PAGE. Red bars indicate CDEI.

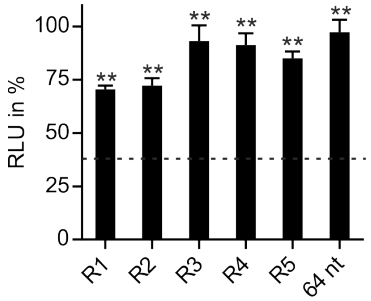

Figure 5.7: Gene regulation is independent of the high AU content of the *UCP3* wt element. Luciferase activity of randomized sequences with the same nucleotide composition (R1-R5) as the *UCP3* wt element and one randomized sequence with uniform nucleotide content (64 nt). Firefly luciferase activity was normalized to *Renilla* luciferase as internal transfection control. Values are normalized to an empty vector control. n = 3. (**) P-value <0.01. Adopted from [113].

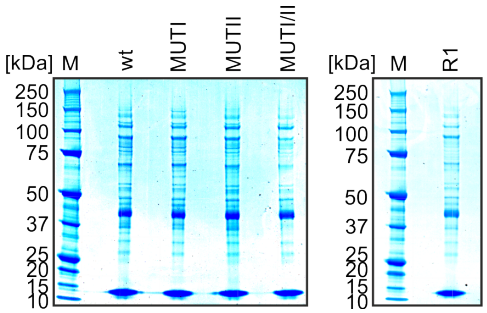

Figure 5.8: Identification of proteins associated with the *UCP3* wt element. Analysis of protein binding to the *UCP3* RNAs shown in (**Tab. 2.1**) and a randomized control with the same nucleotide composition (*R1*). For RNA affinity purification HEK293 whole cell lysates were incubated with the different *UCP3* RNAs. Associated proteins were visualized by SDS-PAGE. M = protein size marker. Adopted and modified from [113].

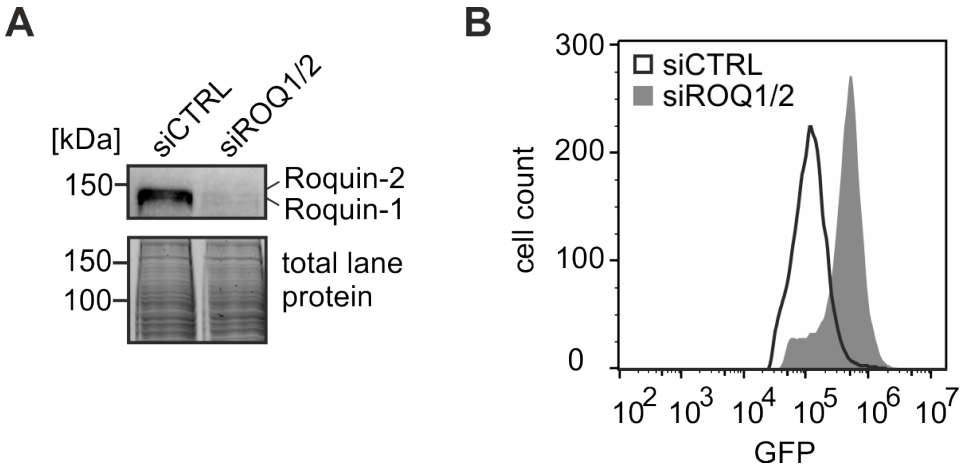

Figure 5.9: Roquin mediates repression by the *UCP3* wt element. (A) Western Blot of Roquin-1 and Roquin-2 after siRNA-mediated knockdown in HeLa cells stably expressing a *GFP* mRNA containing the *UCP3* wt element. Anti-Roquin was used to verify the respective knockdown. Total lane protein is shown as loading control. n = 3. **(B)** GFP fluorescence of *UCP3* wt with and without siRNA-mediated knockdown of Roquin-1 and Roquin-2. GFP fluorescence was measured by flow cytometry. n = 3. Adopted and modified from [113].

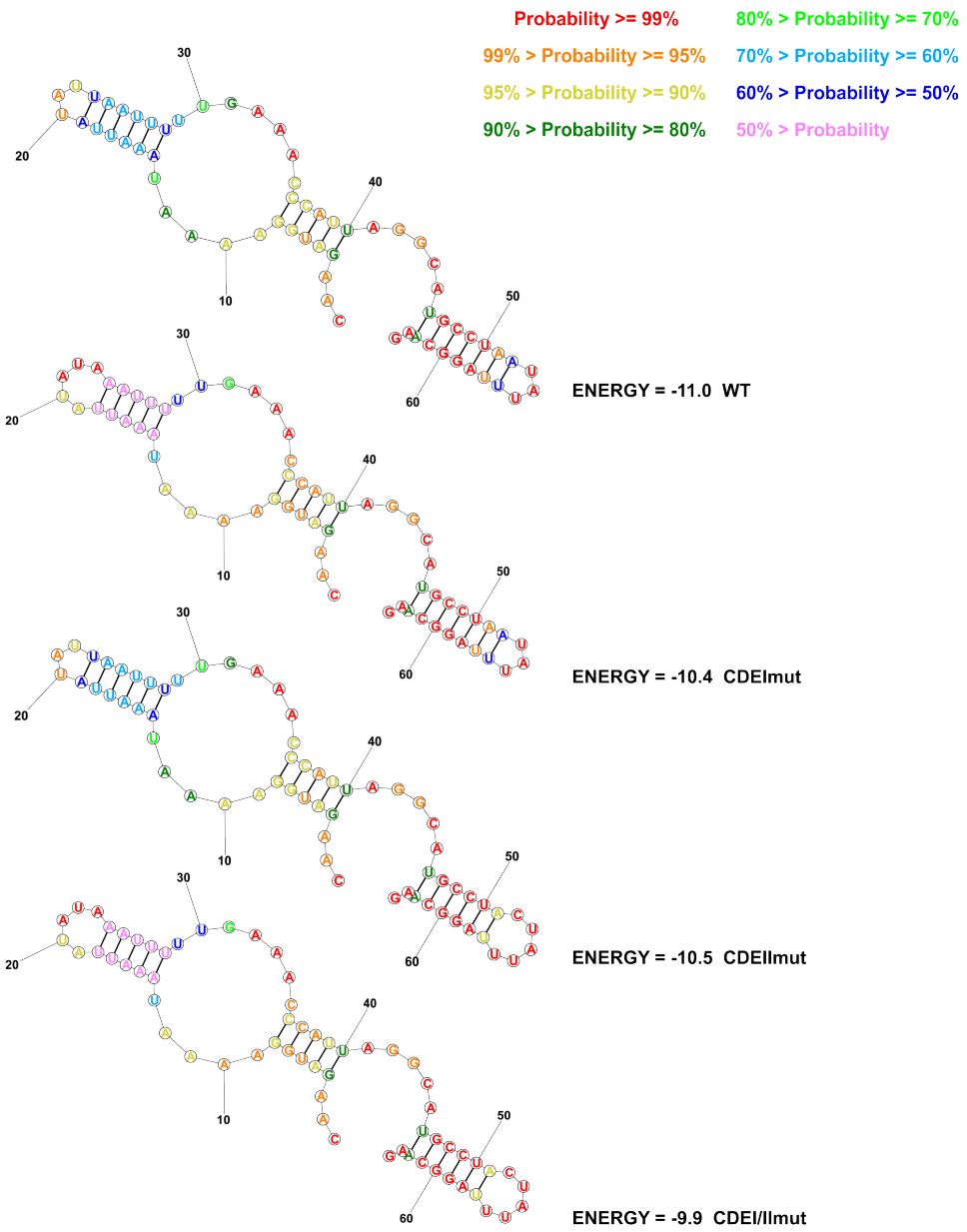

Figure 5.10: Secondary structure prediction of *UCP3* RNAs used for Roquin binding experiments. Secondary structure of *UCP3* constructs used for EMSAs with recombinant Roquin-1 protein predicted by RNAstructure v6.0.1 (`https://rna.urmc.rochester.edu/RNAstructureWeb`). Adopted and modified from [113].

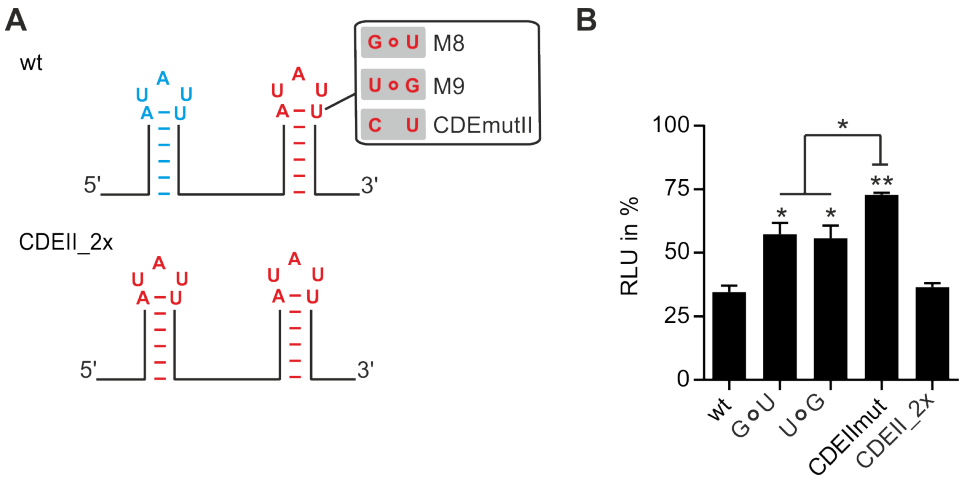

Figure 5.11: Non-canonical closing pairs reduce luciferase expression. (A) Overview of *UCP3* mutants. **(B)** Luciferase activity of *UCP3* mutants. Firefly luciferase activity was normalized to *Renilla* luciferase as internal transfection control. Values are normalized to an empty vector control, without *UCP3* 3'UTR sequences. n = 3. (**) P-value < 0.01. (*) P-value < 0.05. Adopted and modified from [113].

5.2 Supplementary Tables

Table 5.1: Conserved 3'UTR structures predicted by Dynalign and tested for functionality *in vivo*. Adopted and modified from [113].

Gene symbol	Chromosome location	Sequence 5'→3'
ADM	chr11:10306992-10307091	GTGTCACCCCACCAGGGCGCAAGCCTCACTATTACTT GAACTTTCCAAAACCTAAAGAGGAAAAGTGCAATGC GTGTTGTACATACAGAGGTAACTATCA
BCL2L1	chr20:31665060-31665159	CCCCAGAGCTGTTTATGGCCTCAGCTGCCTCACTTCC TACAAGAGCAGCCTGTGGCATCTTTGCCTTGGGCTG CTCCTCATGGTGGGTTCAGGGGACTCA
CTBP1	chr4:1211900-1211999	CATGAACGTTCTTGTCTGTGTACAGTTTTTAGAACAT TACAAAGGATCTGTTTGCTTAGCTGTCAACAAAAAG AAAACCTGAAGGAGCATTTGGAAGTCA
EGLN3	chr14:33925414-33925513	AAATGGAGGTGGTAGATGCCACAGAGAGGCATCACG GAAGCCTTAACAGCAGGAAACAGAGAAATTTGTGTC ATCTGAACAATTTCCAGATGTTCTTAAT
ELK3	chr12:96267231-96267330	CCCACGGGCTAGTTTACCTGTGTCGTGAGAAGGACA TTGTGAAACTCTTGTTAATTTGGTTTGCACTTTTCAT AACATGGATAGTCTAGATTTATGTTAG
EMX1	chr2:72934005-72934104	GGCCCCTGCGTGGGCCCAAGCTGGACTCTGGCCACT CCCTGGCCAGGCTTTGGGGAGGCCTGGAGTCATGGC CCCACAGGGCTTGAAGCCCGGGGCCGCC
HIRA	chr22:19330849-19330948	ATTGAATTTCCTTTGGCCGATAATCAGGATTTCCCTA TAAGTCACTTGGACATTGGTCACTTGTAGGAAATTT AAACTCTAATTATGACAGCTACACTGA
HMOX1	chr22:35393699-35393798	GAAGGCTTTCAGGGCCTCCAGCCCTCTCACTGTGTC CCTCTCTCTGGAAAGGAGGAAGGAGCCTATGGCATC TTCCCCAACGAAAAGCACATCCAGGCAA
HOXC9	chr12:54002875-54002974	TTTAGAGTTAGTTCTACCCAGCGAGGAGGAGGCGGG GAGAGAAACTGCGTTCTCTTTCCCCAGCGCAACCGA AATAAATGACACATACAAATGTGATTTT
MYF6	chr12:80709011-80709110	CCTTCCTGGCCTAATCCTTTAGATTAGGTCACATTAC ATTAACATTTAGGAACCCAGACCGAAAAGTTGCTGA AAGGGAAGGAGACACATTCACAAAGAA
NR4A2	chr2:156325469-156325568	TTTATACCACTGTATTGTGTGTAGTCCATGTTCTAAA TCCAGGATGCCCCGGAGCCAAAATGCCCTTTCAGGT TCTGCCTGCAGGTTAGGAAATAGCAAC
PDE5A	chr4:119498128-119498227	GAATGTTTGGAATCTTTCCTTTTTCAAAAGTAGGTT AGGAGCAAATTATCATACATTCTGTGACATTTAAAG CCTTTATAGGATAGTGAAAAATGCTGGC
RBPJ	chr4:26431008-26431107	CTACCGTCTTTTTGCTAGGACTTAAACTGACTTGAGT GTGGCAAAAAGTTAACAAAAAAGGAGAAAAAATGAA CAATCGTTTGTGGTTTCTTGGGAAAAC
SIX2	chr2:45005909-45006008	GTTCTTGTTTGGGATTTATTTTCAACAAGTTACTTTT AGGATCCTTTTGGGGCTGGAGACTGAGTCTTGAACC ACAGAAGGGAATAAATTATACACCACT
TFDP1	chr13:113641068-113641167	TGTTTACATACTTATAAGTCTATCATTTAAAGACATG TACTGAAACAAATGTATTTGTTTCATAAGCATCTTCC TGTAATCTATTATATAAAATTGAAATTA
TGFBR1	chr9:99153356-99153455	TATTGTATTTGTGCAGGATTCTTTAGGCTTTATCAGT GTAATCTCTGCCTTTTAAGATATGTACAGAAAATGT CCATATAAATTTCCATTGAAGTCGAAT
THBS1	chr15:39595970-39596069	ATGTAAATAGGCACTTAAATAGAAGCAGGAAAGGGA GACAAAGACTGGCTTCTGGACTTCCTCCCTGATCCCC ACCCTTACTCATCACCTGCAGTGGCCA
TLX3	chr5:171311950-171312049	TTTTTTCTTTAGAAACCGGCCACCTGCTTCCCCCGCG GGGGCCGCTGGAGGAAGGGCAGCCGACCCGGCCGCT GGGGGAAGTGCCAGGGGCCCGGGGCAC

Gene symbol	Chromosome location	Sequence 5'→3'
UBE2B	chr5:134390754-134390853	TGTTAAAATACATAACTTCAGTGCAAGAGACTTTGTCACTTATTTCCTTATGTGTGTAGGAGGGGTTAATAAGTCTCTAGCTCTCCATCTATTGATAGTT
UCP3	chr11:74001139-74001238	CTGCCATGCCTCAAGAACACCTTTGTTTTGCACTGACAAGATGGAAAATAAATTATATTAATTTTTGAAACCCATTAGGCATGCCTAATATTTAGGCAAG

Table 5.2: UCP3 tandem CDE sequences and corresponding RLUs. Adopted and modified from [113].

Construct	Sequence 5'→3'	RLU
window	CTGCCATGCCTCAAGAACACCTTTGTTTTGCACTGACAAGATGGAAAATAAATTATATTAATTTTTGAAACCCATTAGGCATGCCTAATATTTAGGCAAG	33.2
miRmut	CTGCCATGCCTCAAGAACACCTTTGTTT--------CAAGATGGAAAATAAATTATATTAATTTTTGAAACCCATTAGGCATGCCTAATATTTAGGCAAG	36.5
del1	----------TTTGTTTTGCACTGACAAGATGGAAAATAAATTATATTAATTTTTGAAACCCATTAGGCATGCCTAATATTTAGGCAAG	32.9
1-7	ACCTCGTGAAAATAAATTATATTAATTTTTGAAACCCATTAGGCATGCCTAATATTTAGGCAAG	34.2
8-14	TCCCCGCAATTATATTAATTTTTGAAACCCATTAGGCATGCCTAATATTTAGGCAAG	55.3
15-21	CAAGATGGAAAATACCGGCGCTTAATTTTTGAAACCCATTAGGCATGCCTAATATTTAGGCAAG	60.0
22-28	CAAGATGGAAAATAAATTATAGGCCGGGTTGAAACCCATTAGGCATGCCTAATATTTAGGCAAG	66.6
29-35	CAAGATGGAAAATAAATTATATTAATTTGGTCCCACCATTAGGCATGCCTAATATTTAGGCAAG	29.9
36-42	CAAGATGGAAAATAAATTATATTAATTTTTGAAACAACGGCTGCATGCCTAATATTTAGGCAAG	40.8
43-49	CAAGATGGAAAATAAATTATATTAATTTTTGAAACCCATTAGTACGTAATAATATTTAGGCAAG	75.1
50-56	CAAGATGGAAAATAAATTATATTAATTTTTGAAACCCATTAGGCATGCCGCCGCGGTAGGCAAG	78.8
57-64	GAAGATGGAAAATAAATTATATTAATTTTTGAAACCCATTAGGCATGCCTAATATTGCTTACCT	80.0
MUTI	CAAGATGGAAAATAAATTATATTAATTTTTGAAACCCATTAGGCATGCCTAATAGGGAGGCAAG	60.1
MUTII	CAAGATGGAAAATACCGGCGCTTAATTTTTGAAACCCATTAGGCATGCCTAATATTTAGGCAAG	73.3
MUTI/II	CAAGATGGAAAATACCGGCGCTTAATTTTTGAAACCCATTAGGCATGCCTAATAGGGAGGCAAG	98.6
CDEImut	CAAGATGGAAAATAAATTATATAAATTTTTGAAACCCATTAGGCATGCCTAATATTTAGGCAAG	66.9
CDEIImut	CAAGATGGAAAATAAATTATATTAATTTTTGAAACCCATTAGGCATGCCTACTATTTAGGCAAG	72.8
CDEI/IImut	CAAGATGGAAAATAAATTATATAAATTTTTGAAACCCATTAGGCATGCCTACTATTTAGGCAAG	88.5
M1	CAAGATGGAAAATAAATTTTATAAATTTTTGAAACCCATTAGGCATGCCTAATATTTAGGCAAG	37.4
M2	CAAGATGGAAAATAAATTGTATCAATTTTTGAAACCCATTAGGCATGCCTAATATTTAGGCAAG	33.4
M3	CAAGATGGAAAATAAATTCTATGAATTTTTGAAACCCATTAGGCATGCCTAATATTTAGGCAAG	38.6
M4	CAAGATGGAAAATAAATTATATTAATTTTTGAAACCCATTAGGCATGCCTATTATATAGGCAAG	40.9
M5	CAAGATGGAAAATAAATTATATTAATTTTTGAAACCCATTAGGCATGCCTACTATGTAGGCAAG	35.2

Construct	Sequence 5'→3'	RLU
M6	CAAGATGGAAAATAAATTATATTAATTTTTGAAACCCATTAGGCATGCCTAGTA TCTAGGCAAG	34.2
M7	CAAGATGGAAAATAAATTGTATCAATTTTTGAAACCCATTAGGCATGCCTAGTA TCTAGGCAAG	36.6
M8	CAAGATGGAAAATAAATTATATTAATTTTTGAAACCCATTAGGCATGCCTAGTA TUTAGGCAAG	57.3
M9	CAAGATGGAAAATAAATTATATTAATTTTTGAAACCCATTAGGCATGCCTAUTA TGTAGGCAAG	55.8
M10	CAAGATGGAAAATAACTTATATTAAGTTTTGAAACCCATTAGGCATGCCTAATA TTTAGGCAAG	32.2
M11	CAAGATGGAAGATACATTGTATCAATGTTTGAAACCCATTAGGCATGCCTAATA TTTAGGCAAG	34.4
M12	CAAGATGGAAAATAAATTATATTAATTTTTGAAACCCATTAGGCATGCGTAATA TTTACGCAAG	33.1
M13	CAAGATGGAAAATAAATTATATTAATTTTTGAAACCCATTAGGCATGCTTAATA TTTAAGCAAG	33.3
M14	CAAGATGGAAAATAAATTATATTAATTTTTGAAACCCATTAGGCATGCCTTATA TTAAGGCAAG	32.5
M15	CAAGATGGAAAATAAATTACATTAATTTTTGAAACCCATTAGGCATGCCTAATA TTTAGGCAAG	41.2
M16	CAAGATGGAAAATAAATTATATTAATTTTTGAAACCCATTAGGCATGCCTAACA TTTAGGCAAG	41.0
M17	CAAGATGGAAAATAAATTACATTAATTTTTGAAACCCATTAGGCATGCCTAACA TTTAGGCAAG	46.3
M18	CAAGATGGAAAATAAATTATGTTAATTTTTGAAACCCATTAGGCATGCCTAATA TTTAGGCAAG	46.8
M19	CAAGATGGAAAATAAATTATATTAATTTTTGAAACCCATTAGGCATGCCTAATG TTTAGGCAAG	51.4
M20	CAAGATGGAAAATAAATTATGTTAATTTTTGAAACCCATTAGGCATGCCTAATG TTTAGGCAAG	61.5
M21	CAAGATGGAAAATAAATTATACTAATTTTTGAAACCCATTAGGCATGCCTAATA TTTAGGCAAG	50.6
M22	CAAGATGGAAAATAAATTATATTAATTTTTGAAACCCATTAGGCATGCCTAATA CTTAGGCAAG	52.1
M23	CAAGATGGAAAATAAATTATACTAATTTTTGAAACCCATTAGGCATGCCTAATA CTTAGGCAAG	73.7
M24	CAAGATGGAAAATAAATTATAATAATTTTTGAAACCCATTAGGCATGCCTAATA TTTAGGCAAG	54.0
M25	CAAGATGGAAAATAAATTATATTAATTTTTGAAACCCATTAGGCATGCCTAATA ATTAGGCAAG	45.6
M26	CAAGATGGAAAATAAATTATAATAATTTTTGAAACCCATTAGGCATGCCTAATA ATTAGGCAAG	75.2
M27	CAAGATGGAAAATAAATTATAGTAATTTTTGAAACCCATTAGGCATGCCTAATA TTTAGGCAAG	51.3
M28	CAAGATGGAAAATAAATTATATTAATTTTTGAAACCCATTAGGCATGCCTAATA GTTAGGCAAG	49.9
M29	CAAGATGGAAAATAAATTATAGTAATTTTTGAAACCCATTAGGCATGCCTAATA GTTAGGCAAG	80.3
M30	CAAGATGGAAAATCAATTATATTAATTTTTGAAACCCATTAGGCATGCCTAATA TTTAGGCAAG	50.2
M31	CAAGATGGAAAATAAATTATATTAATTTTTGAAACCCATTAGGCATCCCTAATA TTTAGGCAAG	73.1
M32	CAAGATGGAAAATCAATTATATTAATTTTTGAAACCCATTAGGCATCCCTAATA TTTAGGCAAG	90.4
R1	ATGAATGTACACTTATAGAAGAATTATGTAGGCTTACTGAACAATACATTAATC TGATGATAAC	70.5
R2	TAAAGAAGACTATAAAAGCGTGAAATATAAGTAGTTTACTCTACCTCAGGAATT ATTTCTTGAA	72.3
R3	TTCGCGATAAAAGAACAGAGAACTAATAGGCATCATAAATTTGATCATGATATA ATTTTAGTTC	93.2
R4	GACGTCTAGACAGCTAAAAAGCGATATTAAAAATGTAGCTTTAAGCTAATTTTA TATCAATGAT	91.3

Construct	Sequence 5'→3'	RLU
R5	GGTGATAATAAAGAATACTTTATCATAACACAGTTTTAGAAGTTTATCGCCGATTGACAAAAAT	85.0
64 nt	GCAAGTGTCCTTCGGTTATCTACGATCTTGACGCAGGCCACTTATGTGTGGGTGCCTACCCCCT	97.2

Table 5.3: Cycle thresholds of miRNAs predicted to bind to the UCP3 element. Adopted and modified from [113].

hsa-miR-	HEK293	Hela
148a	37.4	36.3
148b	36.8	35.9
152	36.9	35.8
130a	n.d.	n.d.
130b	36.6	36.4
301a	n.d.	n.d.
301b	n.d.	n.d.

Table 5.4: High confidence CDEs. Adopted and modified from [113].

Symbol	RefSeqID	chr	start	end
ABCC5	NM_005688	chr3	183638714	183638730
ABHD6	NM_020676	chr3	58280355	58280369
ADAMTS15	NM_139055	chr11	130346460	130346474
ADCY5	NM_001199642	chr3	123001187	123001203
ADGRL2	NM_012302	chr1	82457201	82457215
ADRB1	NM_000684	chr10	115806068	115806082
AGGF1	NM_018046	chr5	76359647	76359661
AKAP10	NM_007202	chr17	19808171	19808187
ANGPTL4	NM_001039667	chr19	8439095	8439111
ARID1B	NM_017519	chr6	157530819	157530833
ARL8B	NM_018184	chr3	5221752	5221766
ARRDC4	NM_183376	chr15	98517021	98517035
ASH1L	NM_018489	chr1	155305857	155305873
ATXN1	NM_000332	chr6	16305397	16305413
ATXN3	NM_001127696	chr14	92527915	92527931
BCAN	NM_198427	chr1	156623183	156623197
BMP2K	NM_017593	chr4	79800482	79800496
BMPR1A	NM_004329	chr10	88683813	88683827
BTAF1	NM_003972	chr10	93788736	93788750
CAMTA1	NM_015215	chr1	7827722	7827736
CASD1	NM_022900	chr7	94186130	94186144
CBL	NM_005188	chr11	119170704	119170718
CCDC117	NM_173510	chr22	29184052	29184066
CCDC47	NM_020198	chr17	61823129	61823145
CCNB1	NM_031966	chr5	68473499	68473513
CD47	NM_198793	chr3	107764478	107764494
CHL1	NM_001253387	chr3	448291	448305
CLK3	NM_001130028	chr15	74922413	74922427
CNOT2	NM_001199302	chr12	70748435	70748449
CNTN3	NM_020872	chr3	74311829	74311845

Symbol	RefSeqID	chr	start	end
CPEB4	NM_030627	chr5	173386101	173386115
CRISPLD2	NM_031476	chr16	84942917	84942931
CSF3	NM_000759	chr17	38173562	38173576
CTDSPL2	NM_016396	chr15	44816836	44816850
CTLA4	NM_005214	chr2	204738615	204738629
CYP1B1	NM_000104	chr2	38295065	38295081
CYR61	NM_001554	chr1	86049170	86049184
DCAF12	NM_015397	chr9	34087944	34087960
DCUN1D1	NM_020640	chr3	182660750	182660766
DDIT4	NM_019058	chr10	74035725	74035739
DDX42	NM_007372	chr17	61896655	61896669
DPY19L1	NM_015283	chr7	34970958	34970974
DSC1	NM_024421	chr18	28710394	28710410
DSCAML1	NM_020693	chr11	117298521	117298537
EFCAB14	NM_014774	chr1	47143920	47143936
EFNA4	NM_182689	chr1	155041915	155041929
EIF3J	NM_003758	chr15	44854207	44854221
EPB41L1	NM_001258329	chr20	34818199	34818213
EPC1	NM_001272004	chr10	32556834	32556852
EPHB1	NM_004441	chr3	134978556	134978570
EPHB3	NM_004443	chr3	184299594	184299608
ESR1	NM_000125	chr6	152424378	152424392
ETAA1	NM_019002	chr2	67637244	67637258
FAM13B	NM_001101800	chr5	137274014	137274030
FAM160B1	NM_020940	chr10	116624238	116624252
FAM43A	NM_153690	chr3	194409617	194409631
FAM46A	NM_017633	chr6	82455479	82455495
FAM8A1	NM_016255	chr6	17608952	17608966
FARSA	NM_004461	chr19	13033456	13033472
FCHSD2	NM_014824	chr11	72549462	72549478
FGF18	NM_003862	chr5	170884516	170884530
FOS	NM_005252	chr14	75748738	75748752
FYTTD1	NM_001011537	chr3	197510494	197510508
FZD8	NM_031866	chr10	35928121	35928137
GABRB3	NM_000814	chr15	26791941	26791957
GID4	NM_024052	chr17	17969423	17969437
GMFB	NM_004124	chr14	54943484	54943500
GREM1	NM_001191322	chr15	33026142	33026156
GRM5	NM_000842	chr11	88238180	88238196
GTF3C2	NM_001035521	chr2	27549237	27549253
HDLBP	NM_001243900	chr2	242166724	242166740
HMX2	NM_005519	chr10	124909746	124909760
HS3ST5	NM_153612	chr6	114378387	114378405
ID1	NM_181353	chr20	30194219	30194233
ID2	NM_002166	chr2	8824387	8824401
IFFO2	NM_001136265	chr1	19232348	19232364
IFFO2	NM_001136265	chr1	19234004	19234020
INSIG1	NM_198336	chr7	155100850	155100864
INSIG2	NM_016133	chr2	118866065	118866079
JADE1	NM_199320	chr4	129795977	129795991
KLF9	NM_001206	chr9	73000093	73000109
KRAS	NM_004985	chr12	25358540	25358556
LAMC1	NM_002293	chr1	183114689	183114703
LHX2	NM_004789	chr9	126795190	126795204
LRAT	NM_004744	chr4	155670433	155670447
MAF	NM_001031804	chr16	79630496	79630512
MAP3K4	NM_005922	chr6	161537948	161537962
MAT2A	NM_005911	chr2	85771791	85771805
MAX	NM_002382	chr14	65541916	65541932
MED1	NM_004774	chr17	37562430	37562446
MFN1	NM_033540	chr3	179110163	179110177
MPHOSPH6	NM_005792	chr16	82182157	82182173

Symbol	RefSeqID	chr	start	end
NCOA1	NM_003743	chr2	24992156	24992170
NFIA	NM_005595	chr1	61923942	61923956
NFIA	NM_005595	chr1	61921872	61921886
NFKBID	NM_139239	chr19	36379227	36379243
NIP7	NM_001199434	chr16	69375861	69375875
NPAS4	NM_178864	chr11	66193461	66193475
NRARP	NM_001004354	chr9	140194952	140194968
NRARP	NM_001004354	chr9	140194260	140194276
NRCAM	NM_001037132	chr7	107788857	107788873
NYAP2	NM_020864	chr2	226516625	226516639
PAIP2	NM_001033112	chr5	138705353	138705367
PARD3B	NM_057177	chr2	206484850	206484864
PCLO	NM_014510	chr7	82449801	82449817
PHACTR2	NM_001100164	chr6	144145245	144145259
PIP5K1A	NM_001135636	chr1	151221686	151221700
PLAGL2	NM_002657	chr20	30783440	30783456
POLD3	NM_006591	chr11	74351877	74351891
PPAN	NM_020230	chr19	10221911	10221925
PPTC7	NM_139283	chr12	110973221	110973237
PROK2	NM_001126128	chr3	71820859	71820875
PTPN4	NM_002830	chr2	120741603	120741617
PTPN4	NM_002830	chr2	120739805	120739819
PTPRT	NM_007050	chr20	40701525	40701541
PWWP2A	NM_001267035	chr5	159503050	159503066
PWWP2A	NM_001130864	chr5	159518937	159518953
RAB5A	NM_004162	chr3	20026498	20026512
RAPGEF2	NM_014247	chr4	160280743	160280757
RAPGEF6	NM_001164386	chr5	130760771	130760787
RET	NM_020975	chr10	43624817	43624831
RFX4	NM_001206691	chr12	107156393	107156407
RIC1	NM_001206557	chr9	5775355	5775369
RSF1	NM_016578	chr11	77377522	77377538
SCMH1	NM_001031694	chr1	41493682	41493698
SEL1L	NM_005065	chr14	81940891	81940907
SEMA6C	NM_001178061	chr1	151104559	151104575
SETD1A	NM_014712	chr16	30995602	30995616
SHC3	NM_016848	chr9	91628200	91628216
SLC25A12	NM_003705	chr2	172641480	172641496
SLC25A37	NM_016612	chr8	23429802	23429816
SLC35F3	NM_173508	chr1	234460006	234460020
SLC35F4	NM_001206920	chr14	58030803	58030819
SLC39A14	NM_001128431	chr8	22278593	22278607
SLC44A5	NM_001130058	chr1	75667928	75667944
SLC9A6	NM_001042537	chrX	135128539	135128553
SNAI1	NM_005985	chr20	48605383	48605397
SNTB2	NM_006750	chr16	69342851	69342865
SOCS2	NM_001270467	chr12	93969847	93969861
SSH1	NM_018984	chr12	109177171	109177187
STRN3	NM_001083893	chr14	31363715	31363731
STXBP5L	NM_014980	chr3	121138907	121138921
TARDBP	NM_007375	chr1	11084178	11084192
TCF7L1	NM_031283	chr2	85536666	85536680
TCF7L2	NM_001146274	chr10	114925879	114925893
TFRC	NM_001128148	chr3	195777829	195777845
THSD7A	NM_015204	chr7	11414601	11414617
TM2D3	NM_025141	chr15	102182566	102182582
TMEM161B	NM_153354	chr5	87491497	87491513
TNFSF15	NM_001204344	chr9	117547098	117547114
TOP1	NM_003286	chr20	39752108	39752122
TRIB1	NM_025195	chr8	126450481	126450495
TRPM7	NM_017672	chr15	50852637	50852653
TSHZ3	NM_020856	chr19	31766076	31766092

Symbol	RefSeqID	chr	start	end
UBE2Q2	NM_001145335	chr15	76193346	76193360
UBR3	NM_172070	chr2	170940511	170940525
UCP3	NM_003356	chr11	73712183	73712199
UCP3	NM_003356	chr11	73712216	73712232
UHRF1BP1	NM_017754	chr6	34845242	34845256
VASH2	NM_001136474	chr1	213162951	213162965
VGLL4	NM_001128219	chr3	11598209	11598225
YAF2	NM_001190977	chr12	42553247	42553263
ZBTB43	NM_001135776	chr9	129600417	129600431
ZC3H12A	NM_025079	chr1	37949442	37949457
ZFPM2	NM_012082	chr8	106816017	106816031
ZSWIM6	NM_020928	chr5	60841392	60841406

Table 5.5: Sequences of the CDEs tested in luciferase assays.

Symbol	Sequence 5'→3'
CYR61	ACAAAAGTAAATGGGAGGGCATTCCATCCCTTCCTGAAGGGGGACACTCC
	ATGAGTGTCTGTGAGAGGCAGCTATCTGCACTCTAAACTGCAAACAGAAA
CYR61_CDEmut	ACAAAAGTAAATGGGAGGGCATTCCATCCCTTCCTGAAGGGGGACACTCC
	ATCAGTGTCTGTGAGAGGCAGCTATCTGCACTCTAAACTGCAAACAGAAA
ZC3H12A	TCTAGCTGTCTGCCTCAGTGGGTCAGAAGCGATCACCCTGTTGATACACA
	TTGTATCTCTGTAGTTTAAGGAGACGCTGCCGGTAACGGCGTCGGTCCGT
ZC3H12A_CDEmut	TCTAGCTGTCTGCCTCAGTGGGTCAGAAGCGATCACCCTGTTGATACACA
	TAGTATCTCTGTAGTTTAAGGAGACGCTGCCGGTAACGGCGTCGGTCCGT
TM2D3	GGAGTTGATGTGGTGTGAGTGATATATTTCTATGTTTTTAATGTACAGCA
	TCTGTACTTTGTTTGCCTTGATAAAGGTAAGATAAATGAAACGCTGAACT
	ATGCTAATCTGGAATTTGTTTTTATTTGCCTGAAATATATTTTTTTCTGT
	GAAAAAATTAAAACGTACTTAAGCCAGGAGAATGAATTATACAGTGATTG
TM2D3_CDEmut	GGAGTTGATGTGGTGTGAGTGATATATTTCTATGTTTTTAATGTACAGCA
	TGTGTACTTTGTTTGCCTTGATAAAGGTAAGATAAATGAAACGCTGAACT
	ATGCTAATCTGGAATTTGTTTTTATTTGCCTGAAATATATTTTTTTCTGT
	CAAAAAATTAAAACGTACTTAAGCCAGGAGAATGAATTATACAGTGATTG
CSF3	TTGAGAGTATCAGGTCTCCCACGTGGGAGACAAGAAATCCCTGTTTAATA
	TATAAACAGCAGTGTTCCCCATCTGGGTCCTTGCACCCCTCACTCTGGCC
CSF3_CDEmut	TTGAGAGTATCAGGTCTCCCACGTGGGAGACAAGAAATCCCTGTTTAATA
	TATAAACAGCAGTGTTCCCCATCTGGGTCCTTGCACCCCTCACTCTGGCC
TNF	AGTTCAGACAGACATGTTTTCTGTGAAAACGGAGCTGAGCT
TNF_CDEmut	AGTTCAGACAGACATGTTTTCTGTCAAAACGGAGCTGAGCT

Table 5.6: High confidence ADEs. Adopted and modified from [113].

symbol	RefSeqID	chr	start	end
NM_018429	BDP1	chr5	70863581	70863598
NM_014856	DENND4B	chr1	153902430	153902447
NM_004338	LDLRAD4	chr18	13646121	13646138
NM_002317	LOX	chr5	121399402	121399419
NM_007013	WWP1	chr8	87479410	87479427
NM_031483	ITCH	chr20	33097856	33097873
NM_000899	KITLG	chr12	88886588	88886605
NM_019008	MIEF1	chr22	39913359	39913376
NM_005923	MAP3K5	chr6	136878232	136878249
NM_020834	HOMEZ	chr14	23743546	23743563
NM_014170	GTPBP8	chr3	112719847	112719864
NM_006843	SDS	chr12	113830675	113830692

symbol	RefSeqID	chr	start	end
NM_173801	OOSP2	chr11	59814873	59814890
NM_000964	RARA	chr17	38513875	38513892
NM_012431	SEMA3E	chr7	82996514	82996531
NM_018027	FRMD4A	chr10	13686433	13686450
NM_016467	ORMDL1	chr2	190635779	190635796
NM_017787	WBP1L	chr10	104575010	104575027
NM_001243812	CACNA1B	chr9	141016502	141016529

Table 5.7: Oligonucleotide sequences used for PCR, reverse transcription and qPCR analysis. Adopted and modified from [113].

Specificity	Sequence 5'→3'
UCP3-fwd	ATAGACCTCGAGTGAACAAGACAAGAAGGCCACTGGTAGCTAAC
UCP3-rev	ATAGACGCGGCCGCGGCATGTGGGGCTGGATCTGGCCCTG
UCP3-del-fwd	GAGGAGGGATTCTGGTCTTCAAGAAAATAAACCAAGATAGATCCATTTG
UCP3-del-ev	CAAATGGATCATCTTGGTTTATTTTCTTGAAGACCAGAATCCCTCCTC
miR152-fwd	GATCTACTCGAGCAGACTCGGCTCCCATCAC
miR152-rev	GATCTACTCGAGCAGACTCGGCTCCCATCAC
pDF-fwd	GATCTAGCTAGCGCACATTTCCCCGAAAAGTGC
pDF-rev	GATCTAGGAGATATCGTCGACAAGCTTATC
pFRT-fwd	GATCTACGATAAGCTTGTCGACGATATCTCCACACACCCGGGAATAAAATATC
pFRT-rev	GATCTAGCTAGCCACTTTTCGGGGAAATGTGC
Luc2_3UTR_seq_fwd	GAGGTGCCTAAAGGACTG
luc2_seq_rev2	AGTGAGCGAGGAAGCTCG
CMV_prom_fwd	TAAGCAGAGCTCGTTTAGTG
SV40_polyA_rev	CACTGCATTCTAGTTGTGG
pHDV_seq_fwd	CACTTTATGCTTCCGGCTCG
AP1_pSP64	TTTAGGTGACACTATAGAATACA
miR130_SL	GTTGGCTCTGGTGCAGGGTCCGAGGTATTCGCACCAGAGCCAAATGCCC
miR148_SL	GTTGGCTCTGGTGCAGGGTCCGAGGTATTCGCACCAGAGCCAACAAAG
miR152_SL	GTTGGCTCTGGTGCAGGGTCCGAGGTATTCGCACCAGAGCCACCAAGT
miR301_SL	GTTGGCTCTGGTGCAGGGTCCGAGGTATTCGCACCAGAGCCAGCTTTG
U48_SL	GTTGGCTCTGGTGCAGGGTCCGAGGTATTCGCACCAGAGCCAACGGTCAG
130a_fwd	CGGCGGCAGTGCAATGTTAAAA
130b_fwd	GGCGGCAGTGCAATGATGAAA
148a_fwd	GCGGTCAGTGCACTACAGAA
148b_fwd	AGCGGTCAGTGCATCACAGAA
152_fwd	GGCGGTCAGTGCATGACAGA
301a_fwd	GGCGGCAGTGCAATAGTATTGT
301b_fwd	GGCGGCAGTGCAATGATATTGT
U48_fwd	GAGTGATGATGACCCCAGGTAA
universal rev-primer	GTGCAGGGTCCGAGGT
GFP_qPCR_fwd1	CGTAAACGGCCACAAGTTCA
GFP_qPCR_rev1	CGCTACCCCGACCACATGAAG
RPLP0_fwd	TCGACAATGGCAGCATCTAC
RPLP0_rev	ATCCGTCTCCACAGACAAGG
DDIT4_F1	GTTTGACCGCTCCACGAG
DDIT4_R1	CATCAGGTTGGCACACAAGT
CYR61_F2	CGAGGTGGAGTTGACGAGA
CYR61_R2	GAGCACTGGGACCATGAAGT
PLAGL2_F1	GCAACCAGAGCAGAGACCAT
PLAGL2_R1	GTCCTTGCGGTGAAACATCT
SNAI1_F1	CTCTTTCCTCGTCAGGAAGC
SNAI2_R1	CGGTGGGGTTGAGGATCT
NRARP_F2	ATGACCAACTGCGAGTTCAA
NRARP_R2	GAACTTGACCAGCAGCTTCA
CSF3_F2	GTCCCACCTTGGACACACTG
CSF3_R2	GACACCTCCAGGAAGCTCTG

Specificity	**Sequence 5'→3'**
PLOD2_F1	GGGGCCAGAAAGTGAGATTA
PLOD2_R1	CCACTTTGTGGTTTGCCTTT
TNFSF15_F2	GCAGACGGAGATAAGCCAAG
TNFSF15_R2	GCCAGGCCTAGTTCATGTTC
ZC3H12A_F1	TCCACTCCCAGAAGAGGAAA
ZC3H12A_R1	GGATGGCACAAACACTGTGA
TM2D3_F1	GGGAAGCCTGTCACTTTTGA
TM2D3_R1	CCAGCAAAATCTGCAAGTCA
CASD1_F1	CATCACCTCCATAGCACCACT
CASD1_R1	TGCAGCTTCATTGTAAGCATCT
ITCH_fwd1	GAAGACGTTTGTGGGTGATT
ITCH_rev1	TAAGAAGCGGGGTTTATCTG
WWP1_fwd1	GACATGGAGATTTTGGGAAA
WWP1_rev1	CAAGGAAAGCTTTGGTCTGT
MAP3K5_fwd2	CAGGAAAACCCCCATTTTAT
MAP3K5_rev2	GCTCTCTTGTCAGGATCTGG
GTPBP8_fwd1	TGCAAATTCTTCACACATTTCT
GTPBP8_rev1	ATGGCTTTAGAGCACCTGAA
hsRC3H1_fwd	TGGACAACCAGAACCACAAA
hsRC3H1_rev	GCTGATCCATTTGGTACATCAC
hsRC3H2_fwd	AAGGTTGGCGCTAATGGTC
hsRC3H2_rev	CAGGAGTCTTGGGTGGAGAA
mmRC3H1_fwd	TTGTACCTGAAGCCACTCAGCAGT
mmRC3H1_rev	TCCACTAGCTGGCAATGAACCAGA
mmRC3H2_fwd	TGCCCATTCTCAGGAAGAGCTTGA
mmRC3H2rev	GCTGTGGTTGTGACAGTGCTGTTT
mmUCP3_fwd	GATGTGGTGAAGGTCCGATT
mmUCP3_rev	GGCATTTCTTGTGATGTTGGG

6 Appendix

6.1 Materials Used in this Study

6.1.1 Disposable Material and Kits

Table 6.1: Disposable material and kits used in this study.

Disposable/Kit	**Company**
T75 culture flask	Greiner Bio-One
Cell culture plates (6-well, 12-well, 24-well)	Greiner Bio-One
Amersham ECL Select Western Blotting Detection Reagent	GE Healthcare
cell culture dishes 60 mm	Greiner Bio-One
Dual Glo Luciferase Assay System	Promega
Dual-Glo® Luciferase Assay System	Promega
Flp-In™ System	Life Technologies
Maxtract tubes (1.5 ml; 2 ml)	QIAgen
MicroAmp® Fast 96-Well reaction plate	Applied Biosystems
MicroAmp™ optical adhesive film	Applied Biosystems
Mini-PROTEAN® TGX Stain-Free™ anyKd gels	Bio-Rad
Nunc™ MicroWell™ white 96-well plates, polystyrene	Thermo Fisher Scientific
PVDF membranes	Bio-Rad
QIAprep® Spin Miniprep Kit	QIAgen
QIAquick Gel Extraction Kit	QIAgen
QIAquick PCR Purification Kit	QIAgen
Reaction tubes (1.5 ml; 2 ml)	Greiner Bio-One
Reaction tubes (15 ml; 50 ml)	Greiner Bio-One
Trans-Blot® Turbo™ Transfer Pack Mini, 0.2 µm PVDF	Bio-Rad
Trypane blue staining kit	Bio-Rad
QIAprep®Spin Miniprep kit	QIAgen

6.1.2 Chemicals and Enzymes

Table 6.2: Chemicals and enzymes used in this study.

Chemical/Enzyme	Company
γ-S-ATP	Biomol
γ-32P-ATP	Hartmann
Acetone	Roth
actinomycine D	Sigma
Agarose	peqLab
Ammonium sulfate	Roth
Ampicillin	Roth
anti Roquin-1/-2 (3F12)	Millipore
anti-rat IgG	Jackson Immunoresearch
ATP	Roth
Biotin-long-arm maleimide	Vector Laboratories
Bovine serum albumin (BSA)	Applichem
Bradford	Bio-Rad
Bromophenol blue	Roth
Calf intestine phosphatase (CIP)	Roche
CIP buffer 10x	Roche
Clorophorm	Roth
Cytidine triphosphate (CTP)	peqLab
Deoxynucleotide triphosphate (dNTP)	peqLab
Dithiothreitol (DTT)	Roth
Dulbecco's Modified Eagle Medium (DMEM)	Sigma Aldrich
Dulbecco's Phosphate buffered Saline (PBS)	Life Technologies
Dynabeads M-280	Invitrogen
EBM-plus	Lonza
ECL select	GE Healthcare
EGM-Plus Singlequots	Lonza
Ethanol	VWR
Ethylenediaminetetraacetic acid (EDTA)	Roth
Fetal Bovine Serum (FBS)	Biochrom
First strand PCR buffer (5x)	Thermo Fisher Scientific
Glycerol, p.a.	Roth
GlycoBlue™	Invitrogen
Guanosine triphosphate (GTP)	peqLab
Heparin	Sigma Aldrich
HindIII-HF	New England Biolabs
Horse serum (HS)	Sigma Aldrich

Chemical/Enzyme	**Company**
hydrochloric acid (HCl)	Roth
Hygromycin B	Invivogen
IGEPAL CA-630	Sigma Aldrich
Isoamyl alcohol	Roth
Isopropylic alcohol, p.a.	VWR
Lipofectamine® 2000	Life Technologies
Magnesiumacetat ($Mg(Ac)_2$)	Roth
MulV reverse transcriptase	Thermo Fischer Scientific
NcoI-HF	New England Biolabs
NheI-HF	New England Biolabs
NotI-HF	New England Biolabs
Opti-MEM®	Life Technologies
Penicillin/ Streptomycin	Thermo Fischer Scientific
Potassium chloride (KCl)	Roth
Proteinase-inhibitor-cocktail (PIC)	Sigma Aldrich
Q5 High-Fidelity DNA Polymerase	New England Biolabs
Q5 reaction buffer (5x)	New England Biolabs
random hexamers	Sigma Aldrich
RiboLock RNase inhibitor	Molox
RNAimax	Life Technologies
RNase T1	Ambion
Roti-Phenol	Roth
Rotiphorese Acrylamide Solution	Roth
Sodium acetate	Roth
Sodium carbonate (Na_2CO_3)	Roth
Sodium chloride (NaCl)	Roth
Sodium dodecyl sulfate (SDS)	Roth
Sodium pyruvate	Thermo Fischer Scientific
Spermidine	Sigma Aldrich
Supersript II reverse transcriptase	Life Technologies
SYBR Green Master Mix (2x)	Thermo Fischer Scientific
SYBR Green PCR Master Mix	Thermo Fischer Scientific
T4 DNA Ligase	New England Biolabs
T4 DNA Ligase	Roche
T4 polynucleotid kinase	New England Biolabs
T4 polynucleotid kinase	Roche
T7 RNA Polymerase	homemade
Taq polymerase	New England Biolabs
TaqMan Fast Universal Master Mix (2x)	Thermo Fischer Scientific
TaqMan® Universal PCR Master Mix	Applied Biosystems
Tetramethylethylendiamin (TEMED)	Roth

Chemical/Enzyme	Company
Thermo Pol buffer	New England Biolabs
Transcription buffer (10x)	Applied Biosystems
Tris	Roth
TRIzol® reagent	Thermo Fischer Scientific
Trypan blue	Bio-Rad
Trypsin/ EDTA	Invitrogen
Turbo Dnase	Life Technologies
TurboDNase buffer (10x)	Life Technologies
Tween® 20	Roth
Universal ProbeLibrary #21	Roche
Urea	Roth
Uridine triphosphate (UTP)	peqLab
XbaI	New England Biolabs
XhoI	New England Biolabs
yeast tRNA	Sigma Aldrich
Zeocine	Invivogen

6.1.3 Technical Equipment

Table 6.3: Technical equipment used in this study.

Technical equipment	Company
Cytoflex S	Beckman Coulter
Centrifuges	Heraeus Christ
ChemiDoc Imaging System	Bio-Rad
Gel documentation with UV screen	INTAS
Infinite® M200 plate reader	Tecan Trading AG
Liquid Scintillation Counter 1409	Wallac
Milli-Q® water purification system with RNase filter	EMD Millipore
Mini-PROTEAN Tetra Cell electophoresis chamber	Bio-Rad
NanoDrop® ND-1000 Spectrophotometer	peqLab
Phosphoimager FL-5000	Fuji
Q-Exactive Plus MS	Thermo Fisher Scientific
StepOnePlus™ Real-Time PCR System	Thermo Fischer Scientific
Storage phosphor Screen BAS_IP MS 2040	Fuji
T100™ Thermal Cycler	Bio-Rad
TC-10™ Automated Cell Counter	Bio-Rad
Thermomixer comfort	Eppendorf AG

Technical equipment	Company
Trans-Blot® Turbo™	Bio-Rad
UltiMate 3000 HPLC	Thermo Fisher Scientific
vacuum slab gel drier	Bio-Rad

6.1.4 Plasmids

Table 6.4: Plasmids generated in this study.

Designator	Plasmid	Section	Description
ADM	pDLP_ADM	2.1.1	
BCL2L1	pLDP_BCL2L1	2.1.1	
CTBP1	pDLP_CTBP1	2.1.1	
EGLN3	pDLP_EGLN3	2.1.1	
ELK3	pDLP_ELK3	2.1.1	
EMX1	pDLP_EMX1	2.1.1	
HIRA	pDLP_HIRA	2.1.1	
HMOX1	pDLP_HMOX1	2.1.1	structually
HOXC1	pDLP_HOXC1	2.1.1	conserved
MYF6	pDLP_MYF6	2.1.1	RNA elements
MR4A2	pDLP_MR4A2	2.1.1	predicted by
PDE5A	pDLP_PDE5A	2.1.1	Dynalign
RBPJ	pDLP_RBPJ	2.1.1	
SIX2	pDLP_SIX2	2.1.1	
TFDP1	pDLP_TFDP1	2.1.1	
TGFBR1	pDLP_TGFBR1	2.1.1	
THBS1	pDLP_THBS1	2.1.1	
TLX3	pDLP_TLX3	2.1.1	
UBE2B	pDLP_UBE2B	2.1.1	
UCP3	pDLP_UCP3delseq	2.1.1	
full 3UTR	pDLP_UCP3	2.1.1	full length UCP3 UUTR
deletion	pDLP_UCP3_DEL	2.1.1	UCP3 3'UTR without predicted element
2x window	pDLP_UCP3_2x	2.1.1	2x UCP3 regulatory element
random ctrl	pDLP_UCP3_RAND	2.1.1	Randomization of the 100 nt element
ΔmiR	pDLP_UCP3mut2	2.1.2	UCP3 element without miR binding sites
miR-152	pCMV_miR152	2.1.2	miR-152 overexpression plasmid

Designator	Plasmid	Section	Description
del1	pDLP_UCP3mut3	2.1.3	
del2	pDLP_UCP3mut5	2.1.3	truncation mutants
del3	pDLP_UCP3mut12	2.1.3	of the 100 nt
del4	pDLP_UCP3mut13	2.1.3	UCP3 element
del5	pDLP_UCP3mut4	2.1.3	
1-7	pDLP_UCP3mut16_1	2.1.4	
8-14	pDLP_UCP3mut16_2	2.1.4	
15-21	pDLP_UCP3mut16_3	2.1.4	1-7 – 57-64
22-28	pDLP_UCP3mut16_4	2.1.4	heptamer mutants
29-36	pDLP_UCP3mut16_5	2.1.4	based on the 64 nt
36-42	pDLP_UCP3mut16_6	2.1.4	del2 construct
43-49	pDLP_UCP3mut16_7	2.1.4	
50-56	pDLP_UCP3mut16_8	2.1.4	
57-64	pDLP_UCP3mut16_9	2.1.4	
MUTII	pDLP_UCP3mut16_3	2.2.1	mutation of CDEI
MUTI	pDLP_UCP3mut7_3	2.2.1	mutation of CDEII
MUTI/II	pDLP_UCP3mut716	2.2.1	mutation of CDEI/II
CDEImut	pDLP_UCP3mut31	2.3.1	mutation of closing bp in CDEI
CDEIImut	pDLP_UCP3mut24	2.3.1	mutation of closing bp in CDEII
CDEI/IImut	pDLP_UCP3mut37	2.3.1	mutation of closing bp in CDEI/II
M1	pDLP_UCP3mut32	2.3.2	
M2	pDLP_UCP3mut46	2.3.2	
M3	pDLP_UCP3mut34	2.3.2	
M4	pDLP_UCP3mut22	2.3.2	
M5	pDLP_UCP3mut25	2.3.2	
M6	pDLP_UCP3mut47	2.3.2	
M7	pDLP_UCP3mut48	2.3.2	
M8	pDLP_UCP3mut55	2.3.2	
M9	pDLP_UCP3mut46	2.3.2	
CDEII2x	pDLP_UCP3mut40	2.3.2	
M10	pDLP_UCP3mut59	2.3.2	
M11	pDLP_UCP3mut15	2.3.2	
M12	pDLP_UCP3mut57	2.3.2	
M13	pDLP_UCP3mut58	2.3.2	M1-M32
M14	pDLP_UCP3mut23	2.3.2	constructs used
M15	pDLP_UCP3mut50	2.3.2	for the mutation
M16	pDLP_UCP3mut51	2.3.2	analysis of the
M17	pDLP_UCP3mut52	2.3.2	del2 construct
M18	pDLP_UCP3mut42	2.3.2	

Designator	Plasmid	Section	Description
M19	pDLP_UCP3mut43	2.3.2	
M20	pDLP_UCP3mut431	2.3.2	
M21	pDLP_UCP3mut35	2.3.2	
M22	pDLP_UCP3mut26	2.3.2	
M23	pDLP_UCP3mut49	2.3.2	
M24	pDLP_UCP3mut62	2.3.2	
M25	pDLP_UCP3mut53	2.3.2	
M26	pDLP_UCP3mut60	2.3.2	
M27	pDLP_UCP3mut63	2.3.2	
M28	pDLP_UCP3mut54	2.3.2	
M29	pDLP_UCP3mut61	2.3.2	
M30	pDLP_CDEI_5bp	2.3.2	
M31	pDLP_CDEII_5bp	2.3.2	
M32	pDLP_CDEI/II_5bp	2.3.2	
R1	pDLP_UCP3R2	2.1.3	R1-R5
R2	pDLP_UCP3R3	2.1.3	randomizations
R3	pDLP_UCP3R4	2.1.3	of del2
R4	pDLP_UCP3R5	2.1.3	
R5	pDLP_UCP3R6	2.1.3	
64 nt	pDLP_64 nt	2.1.3	artifitial sequence with 50% GC content
	pDF_FRT	2.1.5	
UCP3 wt	pDF_FRT_mut5	2.1.5	integration of GFP-UCP3 wt and MUTI/II fusions
MUTI/II	pDF_FRT_716	2.1.5	
TM2D3	pDLP_TM2D3	2.4.2	
TM2D3mut	pDLP_TM2D3mut	2.4.2	
CSF3	pDLP_CSF3	2.4.2	new CDEs and
CSF3mut	pDLP_CSF3mut	2.4.2	corresponding
CYR61	pDLP_CYR61	2.4.2	closing pair mutants
CYR61mut	pDLP_CYr61mut	2.4.2	
ZC3H12A	pDLP_ZC3H12A	2.4.2	
ZC3H12Amut	pDLP_ZC3H12Amut	2.4.2	
TNF	pDLP_TNF	2.4.2	
TNFmut	pDLP_TNFmut	2.4.2	
UCP3	HDV_UCP3pe_N	2.1.4	UCP3 window with T7 promoter and HDV ribozym
wt	pHDV_mut5	2.1.4	de2 with T7 promoter and HDV ribozym
MUTI	pHDV_mut7_3	2.2.1	MUTI with T7 promoter and HDV ribozym

Designator	**Plasmid**	**Section**	**Description**
MUTII	pHDV_mut16_3	2.2.1	MUTII with T7 promoter and HDV ribozym
MUTI/II	pHDV_mut716	2.2.1	MUTI/II with T7 promoter and HDV ribozym
R1	pHDV_R2	2.2.1	R1 with T7 promoter and HDV ribozym
CDEImut	pHDV_mut31	2.3.1	CDEI with T7 promoter and HDV ribozym
CDEIImut	pHDV_mut24	2.3.1	CDEII with T7 promoter and HDV ribozym
CDEI/IImut	pHDV_mut37	2.3.1	CDEI/II with T7 promoter and HDV ribozym